

Towards a new world trade war?

The huge US trade deficit can be reduced in two ways—by fiscal rectitude or protectionism. There is a risk that Congress will seek the second path, which could spell disaster.

THE most remarkable modern economic phenomenon is the financial mess into which the United States has stumbled. For the past six years, since before the beginning of the Reagan administration, the federal government has been running a fiscal deficit which has increased over the years to something like \$220,000 million per annum. Part of the explanation is all too familiar—the unwillingness of the government to cut public spending to match income. More recently, for the past four years, the deficit has been sustained by the Reagan administration's conviction that taxes should be cut in the interests of economic revival, and its determination that they should not even now be increased. The consequence of the deficit has been that the US government has had to borrow money in the international money markets, which has been possible only by making investment in dollar securities attractive, by high interest rates and by the promise that the domestic recovery would be strong and sustained. The consequence of that state of affairs is that the dollar itself has been "strong", which is another way of saying that the dollar has been over-valued relative to most other currencies. And the consequence of that, in turn, is that US manufacturers and farmers have been at a disadvantage compared with those elsewhere in the export markets of the world, whence the large and growing trade deficit, estimated at \$137,000 million in 1984.

This state of affairs has now begun to hurt large sections of the US economy. The plight of the farmers has been clear for the past eighteen months (see *Nature* 14 March, p. 119) but is only partly attributable to the difficulties of selling US agricultural products on the international markets. Steel manufacturers have for longer been under pressure from competitors elsewhere, Europe, Japan and even some developing countries. Now the competition is beginning to hurt the traditionally strong sectors of the US economy, the general run of manufactured goods, and also those in which the United States is technically among the most advanced, electronic and telecommunications equipment for example. The glitter of the economic recovery is being dimmed by the difficulties into which some pockets of the economy have been plunged. The sharp fall in the value of the dollar (by an average of 8 per cent) against other currencies since the middle of last month will help, but only after some time and to a small extent. The fall has not yet been sufficient to correct the whole imbalance of US external trade. So, inevitably and even understandably, the US Congress is under pressure from its constituents to shield them from the unpleasant consequences of the imbalance. Last month's wild talk of a general protective duty on all imports from outside has mercifully been stilled, at least for the time being. Now, the emphasis is on correcting the trade imbalance with Japan, estimated at \$35,000 million in 1984.

Reprisals

The consequences of that are odd, to say the least. On 2 April, a committee of the US Senate passed a bill that would require the federal government to take reprisals against Japan if the trade imbalance does not somehow correct itself. (The bill is not yet a law and, even if it were, its influence would be merely hortatory, but it is a sign of the way the tide of

sentiment is flowing in the United States.) There has also been a flurry of bizarre diplomatic activity, with secret exchanges of special envoys between Tokyo and Washington, ostensibly seeking assurances that the newly deregulated Japanese telecommunications system will be open to equipment made in the United States, but whose political objective has been to demonstrate to the Congress that the administration is not sitting on its hands. (Even if Japan gave up the manufacture of telecommunications equipment and imported everything from the United States, the effect on the US trade deficit would be that of a drop in a bucket; the value of Japanese production in this field last year was a mere \$2,300 million, less than 10 per cent of the total deficit.)

Imbalance

The issue of the trade imbalance with Japan is a red herring, but no less dangerous on that account. By any reasonable test, nobody should be surprised that if the United States has a trade imbalance on the present scale, the largest share of that should accrue to Japan, far and away the industrial economy that is the most successful at producing manufactured goods, from motor cars to electronic consumer products, which are cheap, of high quality and suited to what the international markets need. President Reagan, who is fond of telling European governments that they would be economically more successful if they were more like the United States (which may be true), should spare some time to remind US manufacturers that they should emulate their Japanese competitors.

But is not the competition from Japan unfair? This is the recurring cry, not only in the United States but also in Western Europe. And it is true that some sectors of the Japanese economy are closed to competition from elsewhere. Japanese consumers must pay more for beef than if it were imported from Texas (or Australia), for example, which has been a fierce bone of contention with the United States during the past year. But is the United States saying that the time has come when all governments, itself included, must abandon agricultural protectionism? That would be in the general interest, but no more acceptable politically in Washington than in Tokyo. For the rest, it is far from clear whether the difficulties that US manufacturers report in exporting to Japan are consequences of the supposed conspiracy of the Japanese bureaucracy against imports of any kind. But there would be serious dangers in any sustained campaign by the United States to single out Japan as the target for its chagrin over the size of the trade imbalance, not least of all the risk that the industrial community of the West, established painfully since the Second World War, may be undermined.

Whether or not competition with Japan is unfair is in any case irrelevant to the present difficulty of the United States. And everybody knows what the remedy must be. So long as the federal government's deficit is running at its present level, and while policy remains to bridge the gap by borrowing from overseas rather than by printing money (which would be inflationary), the trade imbalance will persist. So the urgent task for Congress and the administration is to cut the deficit. That is one of the issues with which the present Congress has been faced. Last week, the administration and its party's



leadership were hoping to cut \$52,000 million from next year's deficit. The first thing to be said about even this ambition is that it is far too modest. The second is that it will eventually be for the Congress to decide whether the administration will get the cuts it seeks. The present preoccupation with protectionism, and with the Japanese phenomenon, is a way of hiding from the real need to cut spending or to raise taxes. Painful though these questions are, they will have to be answered at some point down the road on which the United States is travelling dangerously. □

Professional research

How to stimulate research when application is of the essence?

ENGINEERS insist that, while their professional skill is based in science (no longer exclusively physical science), that is only half the story. And similarly, physicians insist that there is more to their craft than applied biology. So much is easily accepted. But in the past several years and especially recently, there has grown up a great sense of confusion about the character and the role of research in both engineering and clinical medicine. In most places, there is a belief that suitably designed research programmes would improve the practice of engineering and medicine, and in some countries there are glowing demonstrations of what these benefits may be. But there is no general understanding of how these few successes may be replicated elsewhere, or how the universities should be most fruitfully involved.

Britain just now has encountered problems on both fronts. Thus the Medical Research Council (MRC) appears to be embarking on yet another reappraisal of its Clinical Research Centre, founded in the 1960s at least partly out of admiration for and envy of the hospital-based research carried out by the US National Institutes of Health at Bethesda (see p. 487). The British experience, unfortunately, has not been nearly as encouraging as people hoped, but for reasons that should have been anticipated.

In one sense, it is bad luck that most of the interest in the past two decades has been in the research by means of which the biological basis of medicine has been so dramatically enlarged, but MRC's clinical centre has also been frustrated in another of its original objectives, that the development of new clinical techniques and therapies should be put on a sounder basis, both by example and by the training of a new generation of research-minded clinicians. In the event, the pace of development in clinical medicine has been determined by people of the kind who have traditionally made the running, usually members of the staffs of teaching hospitals. Most often, in Britain, the costs of these developments have been borne unselfconsciously by the health services. The practical question MRC will soon have to ask is whether its influence on the direction and pace of change in medicine would be greater if more of its resources were spent on research grants, and less on in-house research. The dangers in spending money on research grants, sometimes apparent in the United States, is that the techniques of assessment by peer-review that work reasonably well in basic science are a poor guide to the selection of projects and people in the field of applications. The paper-chasing that sullies much of clinical research in the United States should be a warning of the dangers.

Similar difficulties arise in engineering research. Everybody agrees about the need that basic research should prosper (and in fields such as materials science, British support for the foundation of engineering is sadly deficient). There is also a general conviction that what is called engineering research should also prosper. In the United States, one result has been (see p. 488) that the National Science Foundation has now set up the first six of what is intended as a network of engineering research centres closely linked with universities. The hope is that these will become the well-springs of technological

change in the decades ahead, but that they will also train the executors of innovation. The goal is sensible enough (and the centres closely resemble the network established in Britain over the past decade by the private charity called the Wolfson Foundation, many of whose centres have well served their purpose). But the question remains of what engineering research is like.

The point is nicely illustrated by the dilemma in which the British Science and Engineering Research Council now finds itself. For well over a decade, the council (which is the chief source in Britain of support for academic research) has been under pressure to do more to turn ideas into wealth. Its response has been to spend money on a variety of schemes for inducing collaboration between industry and universities (most of them successful and cost-effective), but also to spend more money on engineering research, whatever that may be. Direct spending in universities and polytechnics in the United Kingdom under the heading of engineering is now more than 40 per cent of the total (which is one reason why spending on, say, basic physics is being skimmed). Yet a committee report, on which the research council is now seeking general opinions, would have the scale of spending further increased, with the objective of turning British higher education into one of the principal sources of technological innovation. An important part of the argument is that the acknowledged differences between science and the practice of engineering require that engineering projects should be directed not at the discovery of new knowledge as such but at the conception and design (not necessarily the development) of new products and machines. The research council is believed to be in two minds about the proposal, and even for the first time to be prepared to contemplate a separate research council for engineering.

This may be the only way to go in a country such as Britain, but everybody should be aware what is entailed. Higher education is for training students, and research (whose first product is scholarship) is also an extension of that process. Most engineering education (like most medical education, especially in Britain) neglects the basic science of the field (which may be part of the reason why British industry has been slow to exploit scientific innovation). There are several conceptual fields, such as the development of techniques for telling complicated computer systems what to do, or in definition of the relationship between the chemical and physical microstructure of a semiconductor and its properties, where it is easy to see armies of talented people providing future engineers with grist for important mills. But if it is to be the role of engineering research within a university environment to take on the precompetitive development of new products, the question must naturally arise of what the role of industrial companies' research and development departments should be. And will it be possible for universities to use their existing techniques of peer-review and the like to decide who should do what, with what resources?

Exploitation

The six new centres established by NSF should yield some first-hand experience of the problems to be solved, although there are already whole institutions such as the Massachusetts Institute of Technology which are partly organised along these lines already. In Britain, the case for following the path that seems now mapped out is that there is a need, unmet by industry, for the more deliberate exploitation of new ideas, that such a development would also lend engineering some of the distinction engineers say it lacks and that there is a need for a stronger relationship between industry and higher education on more general grounds. The arguments against are that universities may turn out to be poor managers of such novel institutions (and worse, that they might persuade academic engineers to follow inappropriate norms of academic attainment). But the essential criterion is that more support for these engineering initiatives in universities should not rob basic research of funds. □

US fusion research

Hope now depends on overseas collaboration

Washington

THWARTED by a probable 17 per cent cut in the US fusion research budget in two years, US scientists are pinning their hopes for further progress on increased collaboration with Europe and Japan. Plans have been abandoned for a US device that would both achieve plasma ignition, the point at which fusion becomes self-sustaining, and simultaneously allow the testing of high-performance technology needed for a commercial fusion reactor.

The US Department of Energy's latest magnetic fusion programme, now before Congress, makes clear that the next US machine will instead concentrate primarily on the physics of plasma ignition; no timetable is specified. Details of how the programme should be implemented are being worked on at Argonne National Laboratory. The best hope now is that, through international agreements now in the final stages of negotiation, US scientists might participate in research on high performance technology on a machine outside the United States.

Waning interest in fusion research in the United States is, according to the Department of Energy, due to the oil glut. The new plan says simply that there is no longer a premium on the early development of fusion power. Instead, the goal now should be to further scientific understanding of different options for plasma confinement. The US fusion research budget fell from \$470 million in 1984 to \$436 million in 1985 and the administration's proposal for the 1986 budget is \$390 million. There seems to be no chance that Congress will increase this amount and, indeed, a further decrease cannot be ruled out despite lobbying efforts by the research community.

The impact of the reduction so far has been felt both in the five national laboratories where fusion research is carried out and in universities. Some job losses now seem inevitable. At the Princeton Tokamak Fusion Test Reactor, plans to introduce tritium (which would be used, along with deuterium, as fuel in a fusion reactor) have been delayed by two years, until 1988, and construction of the MFTF-B tandem mirror device at Lawrence Livermore Laboratory has been indefinitely postponed. The 1970s goal of achieving a commercially feasible fusion reactor by the year 2000 is now acknowledged to have receded by at least 25 years.

The Department of Energy's programme suggests four major areas of research for special attention. These have been agreed in principle with inter-

national partners and will be discussed at the next economic summit meeting at Bonn later in the year. The objectives are the choice of confinement system, the properties of burning plasmas (not now being investigated), the development of suitable construction materials for a future commercial reactor and the development of a blanket for extracting the energy emitted from a fusion reactor in the form of energetic neutrons.

Two international agreements are being negotiated by the Department of Energy under the auspices of the International Energy Agency. These would cover US work on the ASDEX device and the Wendelstein stellarator in West Germany. A further agreement being negotiated with the European Community would formalize US collaboration in the TORE-SUPRA device under construction in France (a steady-state tokamak). Although there have been significant scientific exchanges in the past, legal difficulties—especially over patent rights—

have limited extended stays, according to John Clarke, director of the Department of Energy's fusion programme; separate bilateral agreements with Japan are already in operation.

Fusion researchers are sensitive to the suggestion that they have failed to deliver the goods to their sponsoring agencies: there are encouraging results shortly to be reported from the Princeton tokamak which put the US "well on the way" to achieving the goal of scientific break-even, according to Harold Furth, director of the Princeton Plasma Physics Laboratory. A break-even demonstration is scheduled for 1988. Furth hopes that a small US ignition device could be started in 1988, for completion five years later; experience with low-duty cycle experiments would then start in the mid-1990s. Design work is well advanced on a larger technology-testing device that could be constructed in collaboration with international partners starting in 1990 or 1991. This device, known as INTOR, has been planned by the International Atomic Energy Agency since 1978. If built, experience with high-duty cycle operations could start at the turn of the century. Whether governments will provide the necessary funds remains to be seen.

Tim Beardsley

Clinical research

UK centre under scrutiny again

AFTER one rather futile attempt to help the UK Medical Research Council (MRC) to decide the future of its Clinical Research Centre (CRC), a committee under the chairmanship of Sir Michael Stoker is to try again. The outcome should be an assessment of the first 15 years of the centre's work and is likely to include recommendations as to how CRC should be changed to meet both the changed circumstances in which it finds itself and widespread scepticism of its value.

At a cost of £10.4 million in 1983–84, CRC is the most expensive MRC research establishment. Despite a 20 per cent cut in its running expenses and no money for new equipment last year, CRC consumed nearly nine per cent of the total MRC budget. The cost per unit head is not excessive. The question now is not so much whether there are too many heads—although many would like that to be asked—as whether the centre fulfills its role.

That question can be answered only if the role is clear, but Stoker's first committee seems to have drawn a blank on this point. Nevertheless, it seems agreed that CRC was set up to be a focus and centre of excellence in clinical research in the context of commonplace diseases. That is why it shares a site and facilities with Northwick Park Hospital, an otherwise ordinary district hospital of the Na-

tional Health Service. This arrangement has certain advantages but, according to Sir Christopher Booth, director of CRC for the past seven years and four years from retirement, it also has disadvantages.

One is that there has been little room for medical specialization within the hospital, although the regional health authority now seems willing to change that. A second is that the CRC director has little influence over hospital appointments although, unless those appointed have an interest in research, CRC cannot thrive. Sir Christopher also believes that CRC suffers from not being a recognized centre for postgraduate training.

To some outsiders, including many MRC scientists, the problems of CRC are greater. In 15 years, it has not made its mark as a centre of excellence and has not been quick enough to take aboard new approaches and technology. Why, in particular, is it only now that the application of molecular biology to medicine is beginning to find a place at CRC?

Stoker's second committee has the task of trying again to establish the original aims of CRC but also, and much more importantly, to advise MRC whether the objectives should now be changed and, if so, how to go about it. The committee's report is expected to reach MRC by the end of the year.

Peter Newmark

US engineering research

Six special centres founded

Washington

THE National Science Foundation (NSF) has now designated the first six of its engineering research centres, planned to operate in close association with university departments. They were proposed in 1983 by NSF's engineering directorate in response to widespread concern about the quality of engineering research in US universities. The six centres will be at: the University of California at Santa Barbara; Columbia University, New York; the University of Delaware, with Rutgers University; the University of Maryland, with Harvard University; Massachusetts Institute of Technology; and Purdue University.

are already being devised.

NSF's proposal to designate special engineering research centres was endorsed heartily last year by the National Academy of Engineering, which urged NSF to aim for no fewer than 25 centres. NSF's director, Erich Bloch, said last week that progress towards that goal would depend on the continuing support of Congress. The administration's budget proposal for fiscal year 1986 includes \$25 million for engineering centres, which should allow a further five or six centres to be designated next year.

The subject area to be covered at each centre was proposed by the host institution. NSF did not attempt to identify in

Location	Funds (over five years)	Subject
University of California, Santa Barbara	\$14 million	Robotic systems in microelectronics
Columbia University	\$20 million	Telecommunications
University of Delaware (with Rutgers University)	\$7.5 million	Composites manufacturing
University of Maryland, College Park (with Harvard University)	\$16 million	Systems research
Massachusetts Institute of Technology	\$20 million	Biotechnology process engineering
Purdue University	\$17 million	Intelligent manufacturing systems

The centres will together receive \$94.5 million from NSF over five years, but funds will also be contributed by the host institutions. NSF hopes that its contribution will be matched by state governments and industry. The emphasis is on fostering interdisciplinary research in such areas as robotics, materials science, telecommunications, biotechnology engineering and artificial intelligence; each centre will concentrate on a particular area. Extensive collaboration with industry is expected. The centres will also be used for training, both of graduates and undergraduates; new courses

advance the particular specializations it wanted to support, but merely sought to avoid duplication. The purpose of the centres, according to Bloch, is to "enhance the future competitiveness of US industry in international markets" by furnishing industry with a "solid knowledge base".

At least one US senator had lobbied NSF to locate an engineering research centre in his state, apparently successfully. But Bloch insists that siting decisions had been made exclusively on the basis of peer review of the 142 proposals from 106 institutions. **Tim Beardsley**

NASA budget freeze in prospect?

Washington

THE House of Representatives last week sent a message that it may be in no mood to go along with the Reagan Administration's programme of selective budget increases for science while other non-military portions of the budget are slashed. By a surprisingly large 369-36 vote, the House refused to approve the 5 per cent increase for the National Aeronautics and Space Administration (NASA) that the administration had requested for fiscal year 1986. Instead, if the House gets its way—the Senate has yet to act—NASA's overall budget will be frozen at its present level.

The House offered no prescription for allocating the cut among NASA's many programmes, except to say that \$45 million of the \$375 million reduction should come out of salaries and administrative costs.

The details could be worked out in a House-Senate conference or conceivably left to the NASA administrator.

The President's budget proposal for NASA—which had been approved largely unaltered by the House Science and Technology Committee—specified major increases for the space station (up to \$220 million from the current \$140 million) and for space applications and commercialization (up to \$580 million from \$440 million). The scientific research programmes were due for much more modest rises.

By a much closer margin, the House took a token step towards demanding that the shuttle should pay for itself. NASA had wanted to hold the maximum price for a full shuttle payload to \$71 million, rather than raising prices in 1988. The House ordered the maximum price to increase in 1988 to \$105 million.

Stephen Budiansky

Soviet research

Aleksandrov on self-reliance

SOVIET science must try harder to be self-sufficient, Dr Anatolii P. Aleksandrov, president of the Soviet Academy of Sciences, told the academy's general assembly last month. In some fields in which the Soviets had originally been in the lead, he said, projects had been held up for so long that, in the end, foreign equipment had to be purchased under licence and Soviet work in the field had been terminated. The very scale of Soviet industry seems to be a large part of the difficulty.

Aleksandrov is well known as an advocate of Soviet self-sufficiency. On many occasions, he has somewhat grudgingly referred to international cooperation as necessary in the interests of detente and the elimination of duplicated effort, while emphasizing that the Soviet Union needs no outside help.

Nevertheless, Aleksandrov's new criticisms of Soviet industrial performance are more than his habitual harping on self-sufficiency. His special targets have in common their concern with short-run production.

Thus, said Aleksandrov, the Soviet chemical industry, in spite of its "tremendous" scale, cannot produce the relatively small quantities of reagents needed for high-accuracy analysis. Similarly, in spite of the scale of the nuclear industry, which is capable in theory of supplying the economy with all necessary isotopes and isotope-labelled compounds, many institutes still have to purchase such products from abroad.

Similarly with precision research instruments, although the Soviet instrument-making industry produces a vast quantity of general-purpose equipment each year (to the value of some 30-40 million rubles, including 8 million rubles on systems for the automation of routine analysis), it is not geared to small runs where fewer than 50 items are needed at a time. Moreover, many of the instruments produced are, Aleksandrov said, "frankly speaking, completely useless".

Aleksandrov drew particular attention to the failure, so far, of the Ministry of the Radio Industry to produce the special radar sets needed to monitor soil moisture and salinity. These devices, originally a spin-off from the Venus programme, have an important role to play in the union-wide Food Programme, Leonid Brezhnev's final bequest in 1982 to the Soviet Union. In spite of the efforts of the academy's Institute of Radio Engineering and Electronics in adapting the Venus radars to agricultural use, Soviet industry, Aleksandrov said, has not yet managed to produce the devices so urgently needed. **Vera Rich**

Italian research

Renewed discontent of staff

THE underpaid researchers at the Italian national research council (CNR), the largest state research organization in Italy, are getting impatient. Five months after the hopeful appointment as president of the Milan physiologist, Professor Luigi Rossi Bernardi, they are still awaiting action.

So at least says the Comitato Riceratori, the researchers' committee which now claims 700 of the 2,500 CNR scientists as members. The Comitato appears to be pushing Rossi Bernardi faster than he wishes, or is able, to go. A Comitato spokesman complained last month of Rossi Bernardi's "refusal to make contact with us", and outlined old sores between the CNR staff and the universities, among which pay predominates. Even a top CNR researcher receives the equivalent of only a little over £9,000 a year, net of tax, compared with a top professor's £20,000. The most junior jobs in CNR and among university lecturers are paid about £6,500 and £9,000 respectively.

In Rome last week, Rossi Bernardi said he had now met the Comitato and "we agree on many points". Moreover, he said, "everything they say is obvious—what remains to be done is to do it". But some things are plainly easier to achieve than others.

According to the Comitato, one of the main CNR complaints apart from pay is the relative status of university and CNR staff. This will not be easy to change. The CNR advisory committees which plan the research programmes are dominated by 130 university professors against just 10 CNR researchers. Officially, CNR scientists are described as "collaborators", suggesting a role subservient to the universities. The Comitato also complains that the management of the 270 CNR research institutes and centres falls to university professors so that CNR scientists must "obey decisions without their being consulted". Yet according to statistics prepared by Rossi Bernardi himself, and published in the 1984 CNR report, CNR scientists published six times as many papers per head in 1983 as the average Italian scientist, who published only once in four years.

Thus Comitato members are pressing for at least a career structure comparable with that of the universities. The powerful trades unions, on the other hand, insist on a research "contract" structure for CNR under which researchers would be paid for work done.

The real problem, according to Rossi Bernardi, is to produce a new legal structure for CNR, in which promotion on merit is now virtually impossible. Salaries, taking a quarter of CNR's present

£500 million annual budget, could hardly be increased without such a structure. So will a new law be possible? Rossi Bernardi says he is doing what he can, but that it is "a political problem".

Meanwhile, some reforms are under way. Thus Rossi Bernardi has proposed that CNR researchers should supplement their incomes by contracts and consultancies with industry. Then, acknowledging that CNR "cannot do everything", he says it should concentrate

Planetary missions

Soviet frankness cheers US

Washington

WHAT appears to be the first detailed account of the Soviet planetary exploration programme, including plans for an ambitious mission to Mars and its moons in 1988, has been provided by a delegation of Soviet space scientists in the United States last month. The delegation, led by Dr Valery Barsukov of the Vernadsky Institute, described a plan for soft-landing a small probe on the surface of Phobos, the larger of the two martian moons, and then making it "hop" to other points on the surface by the remote firing of small rockets.

Under the plan, two spacecraft will be launched in 1988, carrying solar-physics packages to study particles and fields in the ecliptic. After entering Mars orbit, one of the craft will assume a circular orbit matching that of Phobos, and manoeuvre to within 50 metres or so of its surface. In collaboration with French scientists, the Soviets are developing a 1-Joule laser that will vapourize and ionize surface material from Phobos, which will then be drawn through an on-board mass spectrometer for analysis.

The second craft will primarily be a back-up that may rendezvous with Demos, Mars's second moon, if the first attempt to land on Phobos is successful.

The delegation also described plans for a lunar polar orbiter to be launched in 1989 or 1990 so as to produce a detailed map of the Moon's geochemistry as well as for the possible mission to Venus in 1991 that would include dropping a lander on to an asteroid, possibly Vesta, on the way.

US scientists who heard the delegation's presentation, given at the Lunar and Planetary Science Conference at Johnson Space Center in Houston, say they were both surprised by the Soviets' frankness and impressed by the ambition of their plans.

Although Soviet scientists had previously outlined their planetary programme (most recently at a meeting arranged by the Planetary Society during

on areas of "strategic" importance.

Rossi Bernardi is also planning to concentrate CNR funds next year on CNR institutes whose performance after assessment is found creditable. He will pay out half the 1986 CNR research funds "on merit" he says, which will be a big change for CNR but one of which the Comitato Riceratori should approve. Indeed, the Comitato says it will agree provided that merit is assessed "by the scientific community rather than by bureaucrats and union representatives". Comitato members "are ready to submit themselves to international competition" said a spokesman. **Robert Walgate**

last year's conference of COSPAR, the Committee on Space Research of the International Council of Scientific Unions, in Graz, Austria), last month's presentation filled in many gaps. It was welcomed as what seemed part of a move away from the secrecy and deliberate vagueness of previous Soviet announcements of future space missions.

Official contacts between the National Aeronautics and Space Administration (NASA) and Soviet space officials were ended in 1982, when the Reagan administration declined to renew a ten-year-old space cooperation agreement, citing the imposition of martial law in Poland. But in recent weeks, different administration officials have revived talk of cooperative space ventures, including a joint rescue training mission involving the US shuttle and the Soviet Salyut spacecraft and coordination of Mars exploration.

Last autumn, President Reagan signed a non-binding Senate resolution sponsored by Senator Spark Matsunaga (Democrat, Hawaii) calling for renewal of the cooperation agreement. Matsunaga has now filed another bill that would set up an "East-West" working group on Mars exploration.

During this hiatus, informal contacts between Soviet and US scientists have continued, typified by the arrangement by which US scientists built an instrument with NASA funds which is being carried on the Soviet Vega mission to Halley's comet (see *Nature* 312, 3; 1985). There have been other informal approaches to US scientists qualified to supply instruments for the Soviet Venus-asteroid and other missions.

A more likely form of low-level cooperation, according to US specialists, is, however, data-sharing and joint calibration of instruments on the Soviet Mars probe and the US Mars Geoscience Climatology Orbiter, due for launch in 1990. The Planetary Society is hoping that concrete plans to these ends will be agreed at the 1986 COSPAR meeting in Toulouse. **Stephen Budiansky**

Soviet espionage

Furore over French allegations

Le Monde, the Paris daily, has caused a furore by publishing what is alleged to be a secret report from the Soviet Ministry of the Aviation Industry on the "utilization of positive foreign experience"—in other words, industrial espionage. The disclosure, which coincided with the opening of high-level Franco-Soviet economic talks, caused the Soviet embassy in Paris to issue denials of this "flagrant disinformation" in even more vigorous terms than usual.

Le Monde claims that the disclosures were leaked to it by French intelligence and were the evidence on which President François Mitterrand expelled 47 Soviet diplomats two years ago. They include a description of the alleged inner workings of VPK, the Commission on Military Industries, which is said to issue "shopping lists" to Soviet representatives abroad, specifying what high-technology information the military establishment needs most urgently.

Claims that such lists exist have surfaced regularly in West Germany, and a recent Swedish survey of alleged Soviet industrial espionage by Charlie Nordblom (*Industri Spionage*, Timbor Förlag, 1984) specifically alleged that the State Committee for Science and Technology and the Academy of Sciences of the USSR are involved in compiling them.

The euphemistic Soviet description of the undertaking is not to be wondered at. At one time, the United States ran a Soviet-watching project, operating from the Weisbaden air force base in West Germany, known as STEP, the Science and Technology Exploitation Program.

The Soviet Union may have had advance warning of the impending disclosures in *Le Monde*. The president of the Academy of Sciences, Dr Anatolii P. Aleksandrov stressed to the general assembly of the academy last month that the Soviet Union intended to rely even less in future on imported know-how, while a few days before *Le Monde* published the documents, Radio Moscow broadcast an extensive interview with an "expert of the Soviet State Planning Committee", Dr Aleksei Petrov.

Petrov complained that the West had long suggested that Soviet technical progress depended on Western know-how, but "facts are facts and they show something quite different". During the past 20 years, Petrov said, the Soviet Union sold to the United States twice as many licenses as it had purchased, and almost two-thirds of Soviet imports consisted of raw materials.

Petrov's data, of course, referred to the civilian sector, while the *Le Monde* documents refer primarily to military aviation. To a Soviet planner, however,

the distinction would probably seem artificial. The extremely low figures for military expenditure published in the Soviet budget (around 18,600 million rubles annually) make sense only if they refer to the replacement of existing military hardware, with military research and development absorbed by the civilian sector.

Vera Rich

Australian veterinary science

Lab. at last

Canberra

AFTER several years of to-ing and fro-ing about its value and safety, the \$A160 million Australian National Animal Health Laboratory (ANAH) was officially opened early this month. The ultra-secure laboratory at Geelong, 70 km south-west of Melbourne, was designed by the Commonwealth Scientific and Industrial Research Organization (CSIRO) for rapid diagnosis of and research into exotic livestock diseases, especially foot and mouth disease virus (FMDV), not detected in Australia since 1872.

Over the years, the unfinished laboratory has been a repeated source of political contention. Two issues have arisen—the high cost of the facility, and the objections of many in Australia to the notion that FMDV might be imported, even if for research only, into a virus-free continent.

The laboratory is designed on the "box within a box" principle, in which the work area of highest security is at the centre of the building and is surrounded by a series of "boxes" of decreasing security classification.

For the present, ANAH scientists are conducting investigations into endemic local diseases such as bovine leukaemia and rhinotracheitis, and are attempting to clone the gene for the two outer protein coats of the local bluetongue virus, a serotype of which was isolated from insects collected in Northern Australia in 1977. In a few months, however, it is likely that work will begin on several exotic diseases including rabies, swine fever virus, Aujeszky's disease (pigs), exotic bluetongue, Newcastle disease (poultry) and vesicular diseases which could be mistaken for FMDV. Ultimately, permission may be granted for work on virulent exotics such as rinderpest or rift valley fever, but the importation of live FMDV has been shelved until 1987 at the earliest.

The laboratory will cost \$A8 million a year to operate, but an outbreak of foot and mouth disease would result in the loss of an estimated \$A3,000 million in export earnings.

Jeffrey Sellar

European research

Curien wants more

Strasbourg

PLANS for using science and technology to improve Europe's competitiveness with the United States were put forward in a speech to a gathering of geologists on 1 April by M. Herbert Curien, minister of research in the government of France. Curien's proposals, which have been hammered out over the past few months by groups of scientists and politicians, are to be presented to the Council of Europe early in May, at a meeting called for by M. François Mitterrand, the French president.

One of Curien's themes was European collaboration in space research and development, chiefly through the European Space Agency (ESA). The past twelve years of ESA's work, Curien said, had been creditable, but he urged last week that by the end of the century, Europe should be in a position not merely to build and launch satellites but to put men and women into space and to provide a full "customer service".

More generally, Curien said, Europe could respond to the apparent superiority of the United States in science and technology only if it were "united, ambitious and autonomous". Specifically, Curien lent his weight to proposals of the European Communities and the European Science Foundation (ESF) for research "networks". Since last September, some 50 proposals have been put forward and have now been whittled down to about a dozen.

The network plan is to involve groups of European research laboratories on common projects, much as several different groups of European geophysicists are now involved in the European Geotraverse project of deep seismic sounding (begun in 1983 by ESF). Last week, Curien referred to neurosciences and polar studies as potentially fruitful areas for collaboration on the same pattern.

Curien also wrung his hands last week over one of Europe's longstanding problems, the difficulty of encouraging mobility within the research community. He said that there were signs, in France, that mobility had recently declined. The difficulty, now as previously, is to devise appropriate incentives for the encouragement of mobility. There remain, said Curien, delays in arranging for the exchange of people, financial assistance to people on the move is lacking while there are even difficulties in transferring equipment across national frontiers. He urged that the impetus for the better coordination of the European research community should come from scientists themselves rather than from rule-bound institutions.

Peter Gambles

Brazilian energy

Physicists seek new nuclear policy

BRAZIL should limit its nuclear power programme to the one US (Westinghouse) reactor now operating, "study" the two West German (Kraftwerk-Union) reactors under construction and take more account of the environmental and social consequences of its grandiose hydroelectric schemes, according to a report prepared by Brazilian physicists and delivered to both houses of the Brazilian parliament late last month.

The physicists are still smarting from their anger at not being consulted when, in 1977, the previous military regime drew up plans with Kraftwerk-Union and others for an eight-reactor grid and a complete nuclear fuel cycle. They now say that the West German deal should effectively be cancelled.

The physicists are thus seizing the opportunity to shift the directions of Brazilian energy policy, now that Brazil has its first civilian elected president for some years, Señor Tancredo Neves. Neves made a cut-back in the nuclear programme part of his election platform—and the idea may have appealed to the electorate partly because the principal nuclear power construction site is at Angra dos Reis, on a beach between São Paulo and Rio de Janeiro in the middle of a popular resort region.

A response to the physicists' proposals has, however, been complicated by Neves's illness (he has had four intestinal operations in the past three weeks) which overtook the 75-year-old president-elect before he or his government could be sworn in. It is likely to be six months before Neves is fit to govern, and in the meanwhile few major policy changes are expected in Brazil.

Nevertheless, Neves has unofficially nominated to the new post of energy minister Dr José Israel Vargas, a physicist who specializes in energy matters and was prominent in Brazil's alcohol fuel (gasohol) programme. Vargas was involved in most of the previous military government's energy projects and is said to understand the economic pressures that led the military to commit Brazil to hydroelectric and nuclear power.

In the mid-1970s, the Brazilian economy in the south-east was growing at up to 12 per cent a year, and the national utility FURNAS sought an 11 per cent annual growth in production. Plans were made for 60 nuclear power plants to produce 75 GW by the year 2000, and construction began of one of the world's largest hydropower complexes on the border with Paraguay.

Brazilian physicists complain that the military's energy plans were made without any democratic consultation, and imply also that there was a hidden defence objective in the agreement with

West Germany.

More sanguinely, a Brazilian government official said that "there was then great euphoria and trust in the future". Now the economy has slumped and atomic energy "doesn't make much sense".

The physicists' report, prepared at a meeting chaired by Dr Luigi Pinguelli Rosa of the University of Rio de Janeiro and attended by the president of the Brazilian physical society, Professor Fernando Souza Barros, and the chairman of São Paulo Electric Power Stations Limited, Professor José Goldenberg, recommends that the new government should:

- Cancel the agreement between the Brazilian nuclear power agency, Nuclearbras, and West Germany for "joint

enterprises".

- Stop the construction of nuclear reactors outside Angra dos Reis, and "study" the continuation of construction of the two German reactors at that site.

- Question the effectiveness of the jet nozzle uranium enrichment supplied by West Germany.

- Restructure the Brazilian nuclear sector to make it more accountable and "civilian".

- Re-examine Brazil's nuclear future by open debate.

- "Give priority" to social and environmental considerations in the planning of future energy projects, particularly nuclear and hydroelectric stations.

- Take dam construction out of the exclusive hands of the electric utilities.

- And take account, in a democratic way, of the effects of dams on local populations and the local environment.

Robert Walgate

Gifts to universities

Threat from US tax reforms?

Washington

PRIVATE universities in the United States, which each year receive close to \$3,000 million in charitable contributions from individuals, have been among the loudest to complain about US Treasury proposals for tax simplification that would reduce the allowable charitable deduction.

Congress has yet to take up tax reform seriously; but already an academic squall has broken out over the extent to which donors are truly influenced by tax considerations.

Under the Treasury proposal, charitable contributions would be deductible against income only to the extent that they exceed 2 per cent of the taxpayer's gross income. In addition, donations of non-cash items, such as real estate, art works and stocks, would receive far less favourable tax treatment. At present, the full appreciated value of such property can be claimed as a tax deduction, but under the proposal, the donor could deduct only the purchase price plus an allowance for inflation or the fair market value—whichever is lower.

According to a study made by Professor Charles Clotfelter of Duke University at the request of the Association of American Universities, overall contributions to universities by individuals would fall to 72 per cent of the present level if the Treasury proposal became law. Yet both the universities and those who challenge predictions such as Clotfelter's agree that the most damaging effect of the Treasury proposal would stem from the proposed lowering of the highest marginal tax rate from the current 50 per cent to 35 per cent. Instead of receiving 50 cents off of his tax bill for each dollar

given, the wealthy taxpayer would receive only 35 cents off. This has placed the universities in an awkward spot: they can hardly be seen to lobby for higher tax rates.

The universities are thus taking the line that the damage inevitably done by lowering rates should at least not be made worse by further restricting charitable deductions. And they are especially concerned about the treatment of non-cash gifts, which account for 40 per cent of their contributions (and a proportionately higher share of contributions from the wealthiest donors). The universities estimate that the Treasury proposals would effectively double the "cost" to donors of such gifts.

A more fundamental question is whether the "cost" of giving matters to donors—particularly wealthy donors. Bruce Davie, chief tax economist of the House Ways and Means Committee, has recently identified an increase of donations by high-income taxpayers over the period 1975 to 1983, despite the reduction of the highest marginal rate from 70 per cent to 50 per cent in 1982.

Clotfelter, in an article to be published shortly, responds that the taxpayers in question had an 86 per cent increase in after-tax income, in constant dollars, during the same period but that their contributions went up by only 23 per cent in constant dollars. The universities are also attempting to prove their case by pointing to a sharp drop in contributions by taxpayers with gross incomes in excess of \$1 million after the lower marginal rate went into effect in 1982; average contributions among this group fell from \$205,000 to \$145,000 in that one year.

Stephen Budiansky

When does life begin?

SIR — The widespread notion “that human life begins at conception” is not based on scientific evidence as J.R. Ling (*Nature* 24 January, p. 262) implied. It results from misleading oversimplifications and misconceptions about the biological basis of human individuality.

That our personal individuality develops after conception is conclusively shown by the natural occurrence of identical twins, triplets and so on. A fertilized egg is not yet “a potential human being” as J.R. Baker (*Nature*, 14 February, p. 524) puts it; left undisturbed, it may instead develop into several human beings, or even more commonly spontaneously abort without producing any. Because the timing of the split of the cell mass that produces identical twins is quite variable, no exact time can be given for the “start” of their individual human bodies — but it is always many days after conception.

If what we are interested in is not the start of individual body form, but the personality of individual humans (surely more important), this develops much later — only after the formation of the nervous system. Consciousness and a sense of self seem to develop only after birth and growing experience of the outside world.

If the vague term “human life” refers not to any of these features of the life of the individual but to human life in general, then this had no clear starting point at all in our gradual evolution from apes.

The notion of giving protection to all potential humans is an argument against contraception that would condemn us to gross overpopulation and misery. What we should protect vigorously are actual individuals possessing human feelings. The consequent legal and ethical necessity to choose a sensible dividing line at which such protection should be given must inevitably be somewhat arbitrary because of the gradualness of development. It is a fallacy to suppose that biology provides an easy answer to the problem.

TOM CAVALIER-SMITH
School of Biological Sciences,
University of London,
King's College,
26-29 Drury Lane,
London WC2B 5RL, UK

SIR — The leading article of 21 February (p. 612) asks “. . . who can object to the proposition that there can be no such thing as a human being without implantation?”

Let us examine how such a conclusion can be reached, beginning with some default assumptions about the meanings of the words used. Human life begins at conception (conception = the beginning of life), commonly taken as occurring at the moment of fertilization, the adjective “human” denoting the fact that both sperm and ovum are of human origin. Human life, in order to proceed, requires

implantation; this also is not disputed. Having proceeded by means of this process, the life in question develops to a greater degree of complexity and, in due course, is (normally) born. At this point we are running into trouble because the same words (“human life”) are used both before and after implantation. We must therefore define our terms carefully, as follows; the words “human being” must refer only to human life which has at some stage been carried in its “mother’s” womb; from this your conclusion follows easily. But perhaps this is not good enough, for as is well known, a baby cannot become a living person unless cared for by adults (if we define our terms carefully), on whom he/she also depends for “emotional and physical security”.

On the other hand, perhaps all this playing around with words is idle sophistry. One could as well say that human embryos are not persons because a person is a human being with full rights under law which a human embryo plainly does not possess. Obviously, it follows that not all human beings are persons.

What is to be gained by all this? The following; the public will become confused, Parliament will produce legislation riddled with loopholes, research will thereby go on and human embryos will continue to be subject to experiment and destruction.

Sir, you do science a disservice.

STEPHEN LOVELL

9 St Barnabas Road,
Cambridge, UK

Sequence data

SIR — A national service for the retrieval and analysis of nucleic acid and protein sequences is being supported by the Medical Research Council (MRC) and is available to academic workers in the United Kingdom. The service is based on 4 databases which are regularly updated: the EMBL Nucleic Acid Sequence Library (currently 2.1 million bases); the GenBank Nucleic Acid Sequence database (currently 3.3 million bases); the Protein Information Resource (0.5 million amino-acid residues) and the NEWAT Library of Doolittle¹. Computer programs are available for database searches² and for the analysis of individual nucleic acid and protein sequences^{3,4}. The Staden programs^{5,6} for the experimental determination of DNA sequence have also been implemented. The service provides documentation of the databases and programs, help-files and facilities for communication within the user community⁷. There are currently some 50 scientists located at 15 sites throughout the United Kingdom accessing the databases.

The service was established in collaboration with scientists from a

number of departments at the University of Cambridge, particularly Dr M. Ashburner and Dr M. Bishop, and with the MRC Laboratory for Molecular Biology. The databases and programs are mounted on the IBM 3081 computer, maintained by the University of Cambridge Computer Service, and there are plans to mount them also on a VAX computer. The databases can be readily accessed through the Joint Academic Network (JANET) or through PSS from any location in the United Kingdom. Those interested should contact me or Dr Geoff Kneale at the address below (telephone: 0223 66499) for further information.

OLGA KENNARD

University Chemical Laboratory,
Lensfield Road,
Cambridge CB2 2EW, UK

1. Kneale, G.G. & Bishop, M.J. *Computer Applications in the Biosciences* 1, 11-17 (1985).
2. Bishop, M.J. & Thompson, E. *Nucleic Acids Res.* 12, 5471-5474 (1984).
3. Staden, R. *Nucleic Acids Res.* 10, 2951 (1982).
4. Bishop, M.J. *Biochem. Soc. Trans.* 12, 1008-1010 (1985).
5. Staden, R. *Nucleic Acids Res.* 12, 499-503 (1984).
6. Staden, R. *Nucleic Acids Res.* 10, 4731-4751 (1982).
7. Kneale, G.G. & Kennard, O. *Biochem. Soc. Trans.* 12, 1011-1014 (1985).

Rings of light

SIR — Readers of *Nature* who look for intriguing optical phenomena in the world around them will have enjoyed Professor Thomas Gold's observation of puzzling reflections in water¹. He will, I am sure, regard it as a shared enthusiasm, and not a criticism, if I point out that others have observed and recorded these twists and loops too.

Minnaert's book² contains at section 14 (“Irregular reflection on slightly rippling water”) a general bibliography stretching back to 1840, and he says of his own Fig. 17 “Remarkable is the appearance of closed coils of light . . .”

The most comprehensive discussion is perhaps in a book by Sir Montagu Pollock³ published in 1903. His chapter on “Reflections in rippled water” outlines the geometry, though more particularly of the reflection of horizontal spars rather than of vertical masts. His book is illustrated with many plates by the Yorkshire marine photographers F.M. Sutcliffe and C.E. Wanless, who were fascinated by just such reflections of ships seen in the waters of Whitby and Scarborough harbours: Sutcliffe's Plate XXV is remarkably similar to Professor Gold's, and his Plate XXXV is appropriately titled by the artist “Rings”.

D. W. DEWHIRST

Institute of Astronomy,
The Observatories,
Madingley Road,
Cambridge CB3 0HA, UK

1. Gold, T. *Nature* 314, 12 (1985) and cover photograph, 7 March issue.
2. Minnaert, M. *Light and Colour in the Open Air* (Bell, London, 1940; reprinted by Dover, London).
3. Pollock, M. *Light and Water, a Study of Reflexion and Colour in River, Lake and Sea*, (Bell, London, 1903).

Monitoring quantum jumps

A neat theoretical calculation shows how to monitor the state of a simple model atom without interfering with its quantum state. But Heisenberg's principle remains intact.

IF Heisenberg's uncertainty principle means anything at all, it must surely mean that it is impossible to keep a continuous record of the transition of a quantum system, say an atom, between two distinct quantum states. That is the general expectation. Suppose, for the sake of argument, that a particular atom may be found in one of only two states, a ground state and an excited state, and that some measurement is carried out to establish which state is occupied. The result is certain to be that the measurement will interfere with the system. Ordinarily the atom will be left in the condition that the measurement reveals, at least until something happens to shift it into the other state. So an ambition to monitor the state of the atom continuously would seem to conflict with these familiar expectations.

But, it seems the outlook may not be that unpromising. Richard J. Cook of the United States Air Force Institute of Technology at Dayton, Ohio, and H.J. Kimble of the University of Texas at Austin, have devised what seems to be a neat way around the difficulty (*Phys. Rev. Lett.* **54**, 1023; 1985). Their proposal hangs on the techniques by means of which it has now become possible to confine single atoms (strictly, ions) in a small region of space by means of electromagnetic traps of various kinds (of which the Penning trap is the prototype). Spectroscopists nurse ambitions to use such traps as ways of carrying out spectroscopy with single atoms. Cook and Kimble's proposal for the direct observation of quantum jumps is one such application. To be fair, the authors do cautiously use the word "possibility" in the title of their paper. They go on to add that the effect which they predict has not yet been observed, "but probably will be in the near future".

This is how the proposed Cook/Kimble experiments would work. The first need is for an atom with two excited states which are connected to the ground state by transitions that are respectively strong and weak. Put one such atom into a Penning-like trap (which is naturally easier said than done) and illuminate it with a laser with a frequency which corresponds to the excitation energy between the ground state and the strongly coupled excited state, and measure the intensity of the fluorescence radiation of the same frequency which is given off by the single atom. What happens is merely that the atom is repeatedly excited

(by the laser beam) and de-excited (by spontaneous radiation emission). If the ground and excited states are indeed strongly coupled, the rate of emission of fluorescence radiation can be as much as 100 million photons a second, well within the capacity of radiation detectors. And while, as Cook and Kimble say, the fluorescent signal will in reality be a succession of quantum events, there will be no difficulty in recording this as steady current in, say, a photomultiplier system. That is the first step.

Next, take a laser whose frequency corresponds to the energy difference between the ground state and the other excited state, that to which the ground state is only weakly coupled, and simultaneously illuminate the single atom with that as well. The predominant process will be the repeated excitation and de-excitation of the strongly coupled state, but occasionally the weak excitation will be excited instead. And when that happens, the electron whose oscillation into and out of the strongly coupled state is responsible for the main fluorescence signal will be out of action, marooned in the weakly coupled state instead. And since the coupling is weak, the main fluorescence signal will be turned off for the length of time the electron stays in the weakly coupled state. So the output from the photomultiplier circuit will consist of a series of signals corresponding to the full value of the strong fluorescence punctuated by gaps when the signal vanishes, and which correspond to the times when the weakly excited state is occupied. That, then, is the running record of when the weakly excited state is occupied. The signal from the photomultiplier will be, for all practical purposes, like telegraph signals in which the dots and dashes have random length. There will be no difficulty in interpreting the record with macroscopic measuring devices — the output signal can be drawn on a moving paper chart and interpreted by eye, if need be.

Cook and Kimble attribute the origin of their neat proposal to H. Dehmelt, who is said to have suggested a scheme rather like this in the context of the detection and measurement of weakly coupled optical transitions. Their own chief purpose is to provide a theoretical calculation of the details of the effects of competitive excitation within their three-level model atom on the fluorescent signal eventually produced, and in particular to calculate the statistical

properties of the gaps in the fluorescence signal in terms of the parameters describing the two transitions.

The argument follows a now familiar line. The rate of excitation of each state is determined by the Einstein parameters from each transition, that which determines the rate of spontaneous emission (conventionally labelled *A*) and that which determines (with the intensity of the exciting radiation) the rate of excitation (labelled *B*). Cook and Kimble simplify the algebra somewhat by supposing that the light exciting the strongly coupled transition is intense enough to saturate the single atom (so that, if the weakly coupled excitation did not exist, the atom would spend exactly half of its time in the ground state and the excited states). The practical problem is merely to calculate the probability that the fluorescence signal due to the strong transition will persist for a length of time, say *T*. Inevitably, the results all depend on the Einstein parameters for each transition, but their practical value is that they involve the Einstein *A* parameter representing the rate of spontaneous de-excitation of the weakly-coupled state, a quantity that is not easily determined by conventional measurements.

This neat recipe for an experiment not yet carried out will obviously spur on the single-atom spectroscopists. On the assumption that the result is that which has now been predicted, the question obviously arises of the sense in which Heisenberg's uncertainty principle must be qualified. Fortunately, a little reflection will show that the question is a red herring. The underlying assumption in what Cook and Kimble have done is that the rate of transition between the ground state and the strongly coupled excited state is very much greater than the rate of transition to the weakly-coupled state. Moreover, there is no way in which this system could be used for arranging that the atom is excited one way or the other to order, for the state of the atom at any time is determined by strictly random processes.

So despite this apparent assault, the uncertainty principle remains intact. What does, however, emerge from this neat piece of experimental design is a vivid illustration of the interesting things that will be possible when single-atom spectroscopy moves into the laboratory, which should not be much delayed.

John Maddox

Cellular immunology

Immunological help at last

from Jonathan C. Howard

AN opportunity to celebrate the end of nearly twenty years of ignorance about an essential immunological relationship is provided by the paper of Lanzavecchia on page 537 of this issue¹. The relationship is that which forms between T cells, B cells and antigen molecules when T cells provide help to B cells during the induction of antibody formation. The conceptual significance for cellular immunology of what has now been accomplished is difficult to overstate since T-cell help provides the paradigm for all specific regulation of the immune response. Perhaps the best evidence that something important has happened recently is that the two best up-to-date textbooks^{2,3} contain no hint of the elegant and, I will risk saying, certainly correct model that I am about to describe.

In Burnet's clonal selection theory, each clone of lymphocytes was autonomous, distinguished from others by the specificity of its cell-surface receptor for antigen. In theory, the selecting power of antigen alone was sufficient to define lymphocyte clones destined for antibody synthesis. However, the 'autonomous' model of antibody induction had to be abandoned as a result of pressure from two directions. First, Burnet's 'lymphocytes' were shown to consist of two developmentally independent and functionally distinct lineages of cells. In particular, thymus-derived (T) lymphocytes never turned into antibody-producing cells, whereas marrow-derived (B) lymphocytes did so, but only if T lymphocytes were also present. T-helper cells therefore somehow helped B cells to make antibody. Second, it became clear that T cells made an antigen-specific contribution to B-cell induction from studies on the structural properties required for antigens to be immunogenic. For an antibody to be raised in animals against any defined small determinant or hapten it had to be attached to a carrier molecule that was an immunogen in its own right. In a series of influential experiments, Mitchison⁴ showed that the requirement for T-helper cell in the induction of antibody responses against hapten-carrier conjugates resulted from the obligatory recognition by T cells of antigenic determinants on the carrier. In addition, hapten and carrier had to be linked on the same molecule. To accommodate these facts, Mitchison proposed the antigen-bridge model for T-cell help (Fig. 1), in which the antigen is bound through a haptenic determinant to the B-cell receptor, which is immunoglobulin (Ig) and is simultaneously bound through a carrier determinant to the T-cell receptor. For over a decade this model has provided the only accessible mental image of antigen-specific T-B collaboration.

It was 'MHC restriction' that ultimately destroyed the antigen-bridge model. It is an operating principle of the immune system that T cells can recognize and be activated by antigen only when the antigen is associated with a molecule of the major histocompatibility complex (MHC) on a cell surface. It was shown first by Katz *et al.*⁵ (in a remarkable experiment that anticipated by quite some time the conceptual advances needed to explain it fully), and subsequently by others, that antigen-specific T-B interactions are MHC restricted in just this sense. To visualize the structural basis of such interactions, it was supposed that the antigen specific T-cell receptor forms an antigen bridge with the B cell, as in the Mitchison model, while a hypothetical second T-cell receptor binds independently to MHC molecules expressed

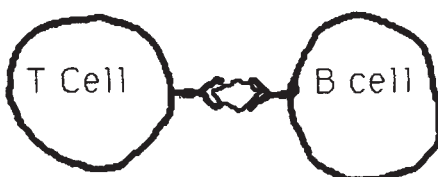
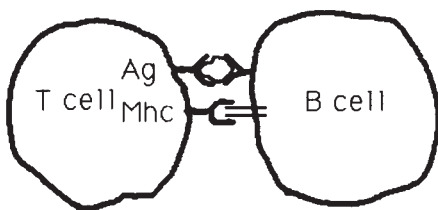


Fig. 1 Antigen-bridge model. The T cell binds to a carrier determinant on the antigen, the B cell binds to a distinct haptenic determinant.

Fig. 2 Modified antigen-bridge model, incorporating MHC restriction. The T cell is given a second receptor (Mhc) with which to bind B-cell Ia molecules. The T-cell receptor for the antigen (Ag) is the same as in Fig. 1. Sometimes the two independent binding sites for Ia and antigen are shown as attached to the T cell via a single stem.



on the B-cell surface. It is an interesting commentary on the importance of the visual imagination that this dual-receptor image (Fig. 2) should have persisted at a time of general acceptance that MHC-restricted T-cell recognition of antigen is mediated by a single receptor with a single combining site⁶. It has, however, been difficult to see how, with a single receptor, a T-helper cell could first be activated by antigen associated with an MHC molecule on a presenting cell, and then deliver help in an MHC-restricted manner to a B cell bearing the antigen in its Ig receptor. For a single receptor, how could the structure antigen + MHC be equivalent to the structure antigen + Ig + MHC?

The resolution of this paradox began when Benacerraf⁷ pointed to strong

circumstantial evidence that protein carriers were not normally in their native configuration when they were recognized in an MHC-restricted manner by T-helper cells. Both he and Unanue⁸ suggested — and it was later demonstrated⁹ — that native antigen is taken up by phagocytic cells, partially degraded and then returned to the cell surface as peptide fragments in association with MHC class II molecules (H-2 Ia in the mouse, HLA-DR in the human). This complex then stimulates specific T-helper cells. Similar peptide/Ia associations could exist or B cells and be recognized by active T-helper cells. Furthermore, the Ig receptor of B cells should concentrate specific antigen on some B cells rather than others.

The clearest early formulation of the new model that I know is from Uhr and his colleagues¹⁰. If T-helper cells recognize Ia-associated fragments of carrier antigens presented on macrophage surfaces, they wrote in 1980, "this presumption implies that the B cell can only be specifically helped by a T-helper cell if a similar antigenic fragment is presented by the B cell. Thus, we could hypothesize that the B cell can pinocytose and degrade antigen like the macrophage. The only difference would be that Ig receptors for specific native antigen equip the corresponding B-cell clone with a means for capturing antigen which is then processed and presented by the B-cells' Ia antigens to the T-helper cell."

This new model of antigen-specific help, represented in Fig. 3, has been substantiated by some particularly striking and elegant experiments, of which Lanzavecchia's, on page 537, represent the culmination. It has been crucial to show that the B-cell Ig receptor functions as an antigen-concentrating device, rather than an antigen-presenting device, for specific B cells. Grey and colleagues were the first to show that molecules captured by B-cell Ig receptors can be efficiently presented to T cells¹¹. They went on to show that B-cell lymphomas can internalize and process non-specifically-bound antigen, and re-present it in the form of peptide fragments to MHC-restricted T-helper cells¹².

These important experiments fell short of demonstrating that specific antigens captured by the B-cell Ig receptor normally are processed and re-presented in that way. Last year, Rock *et al.*¹³ found that TNP-specific B cells, purified from an immune spleen, could present a TNP-carrier conjugate to carrier-specific helper T cells; a strikingly low concentration of conjugate was required compared with that required for presentation by unselected normal B cells. These results showed, for the first time, that specific antigens captured by the Ig receptor are ultimately presented very efficiently to T cells: what they did not prove is that antigens thus captured are processed before presentation.

Lanzavecchia has finally settled the issue using antigen-specific T-cell and B-cell clones from the same individual and prepared against the same antigen, tetanus

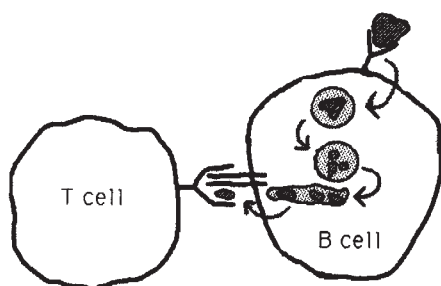


Fig. 3 The new model. The T cell only has one receptor, which recognizes a complex of processed antigen and Ia molecule on the B-cell surface. Antigen processing is initiated by attachment of native antigen to the Ig receptor.

toxoid (TT). The B-cell clones, transformed by Epstein-Barr virus, possess Ig receptors with affinity for TT and present TT in an MHC-restricted manner to TT-specific T-cell clones. In terms of free antigen concentration, TT capture by the B-cell Ig receptor is approximately four orders of magnitude more efficient than non-specific uptake of TT. Lanzavecchia shows that the Ig receptor is involved only with antigen capture, and not directly with antigen presentation, by separating the two events in time. B cells are pulsed briefly with antigen, washed and then allowed to present the antigen to T cells: while the uptake of antigen during pulsing can be prevented by an antibody to Ig, the same reagent has no effect on the ability of pulsed B cells to present antigen to T cells.

As the new model requires, uptake of the TT is not enough for its presentation to T-helper cells; the antigen must also be internalized and processed. Thus, uptake is followed by a stage which is sensitive to the lysosomotropic agent, chloroquine, already known to prevent antigen processing¹⁴. After this stage, antigen-pulsed B cells can be fixed with glutaraldehyde and still present antigen to T cells. By analogy with other work¹², the fixative-resistant state occurs when processed derivatives of the antigen have been returned to the cell surface and have become associated with class II MHC molecules.

Although the haptenic determinant plays no necessary role at the time when help is delivered to the B cell, and although antigens are digested into peptides before presentation, the requirement of hapten-carrier linkage is satisfied by the new model. Only B cells with a receptor for determinants on the native molecule can present carrier determinants efficiently to T cells.

As long as antigen-bridge models were current it was assumed that if the B cell recognized one part of the molecule, then the T cell must recognize another, so that both could bind together to form the antigen bridge. In the new model this condition disappears completely. B cells of any specificity that can bind and internalize an antigen can presumably receive help from any T cell that can recognize the processed form of the molecule.

Like Grey and Rock, Lanzavecchia

assays the response of T cells to antigen presented by B cells. For help, it is clearly the response of the B cells to T-cell recognition that counts. Lanzavecchia has recently shown (personal communication) that EBV-transformed B-cell clones respond to antigen-specific interaction with T cells. In the same antigen dose range that stimulates T cells, specific T-cell clones induce what seems to be terminal differentiation in the EBV clones; cell division is arrested and immunoglobulin release increases.

The quantitative implications of the new findings give as much pleasure as the qualitative. Both Rock *et al.* and Lanzavecchia show that B cells can accumulate and present immunologically significant antigen with an external concentration in the range of 10^{-12} M, which is low enough to be physiologically meaningful. In Lanzavecchia's case, the affinity of the B-cell receptor for TT is about 10^{-8} l mol⁻¹. At an external concentration of 10^{-12} M, the mean occupancy of the B-cell receptor by a univalent ligand such as TT would be about 0.01 per cent, or about 10 molecules per B cell with 10^5 Ig receptors. Surely this is not enough to process and present to T cells. An attractive solution to this paradox is that continuous receptor recycling allows specific B cells to accumulate processed material. Clearly, this option, which I call 'pumping', is not available to B cells in an antigen-bridge model. In confirmation of the pumping hypothesis, Lanzavecchia has now shown (personal communication) that it takes longer for B cells to acquire the ability to stimulate T cells, the lower the external concentration of antigen.

Lanzavecchia's experiments are a considerable achievement. The clarity and elegance that come from studying the interaction between two antigen-specific clones are bought at a high price in technical skill. Nor are there many ex-

periments in cellular immunology of such great pedagogic value as these, which display so accurately the formalism that they are intended to represent. It is a special bonus that these experiments are with human, not rodent, cells.

I do not doubt that the new model is essentially correct. There are however, unresolved problems of considerable interest. Does the binding of antigen to the B-cell receptor deliver a necessary differentiative signal to the B cell, or does the B-cell receptor function only to capture antigen? How is the recognition of processed antigen-MHC complexes transduced into a differentiative signal for the B cell (or, indeed, for the T cell) and what is the causal relationship between this signal and isotype-switching? To what extent does B-cell proliferation and differentiation depend on the availability of non-specific (in an immunological sense) growth factors of either T-cell or B-cell origin? These are important questions but none, surely, so central to immunology as the one that has finally been settled. □

1. Lanzavecchia, A. *Nature* **314**, 537 (1985).
2. Paul, W.E. (Ed.) *Fundamental Immunology* (Raven, New York, 1984).
3. Hood, L.E., Weissman, I.L., Wilson, J.H. & Wood W.B. *Immunology* 2nd edn (Benjamin-Cummings, 1984).
4. Mitchison, N.A. *Eur. J. Immun.* **1**, 18 (1971).
5. Katz, D.H., Hamaoka, T., Dorf, M.E., Maurer, P.H. & Benacerraf, B. *J. exp. Med.* **138**, 734 (1973).
6. Matzinger, P. *Nature* **292**, 497 (1981).
7. Benacerraf, B. *J. Immun.* **120**, 1809 (1978).
8. Unanue, E. *Adv. Immun.* **31**, 1 (1981).
9. Unanue, E. & Ziegler, K. *J. Immun.* **127**, 1869 (1981).
10. Lipscomb, M.F., Lyons, C.F., Ben-Sasson, S.Z., Tucker, T.F. & Uhr, J.W. in *Macrophage Regulation of Immunity* (eds Unanue, E. & Rosenthal, A.) (Academic, London, 1980).
11. Chesnut, R.W. & Grey, H.M. *J. Immun.* **126**, 1075 (1984).
12. Shimonkevitz, R., Kappler, J.W., Marrack, P. & Grey, H.M. *J. exp. Med.* **158**, 303 (1983).
13. Rock, K.L., Benacerraf, B. & Abbas, A.K. *J. exp. Med.* **160**, 1102 (1984).
14. Ziegler, K. & Unanue, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 175 (1982).

Jonathan C. Howard is at the Agricultural and Food Research Council Institute of Animal Physiology, Babraham, Cambridge CB2 4AT, UK.

Gamma-ray astronomy

How much H₂ in the Milky Way?

from Leo Blitz

ON PAGE 511 of this issue Bhat *et al.* deliver a new salvo in a battle that has been raging among astrophysicists since 1980 over how much molecular hydrogen there is in the Milky Way. The crux of the issue is how to convert observed line strengths of the CO molecule into absolute quantities of molecular hydrogen; Bhat *et al.* argue for a lower quantity than any yet published.

The debate has been fierce, emotional and at times has degenerated into *ad hominem* attacks. But why so much bile over such a seemingly esoteric quantity? The reason is that the real issues are much larger astrophysical questions such as: how, when and why do stars form in a spiral galaxy; why do spirals look different from one another; and how do these spiral

galaxies evolve?

The source of the controversy dates to the years around 1970 when millimetre-wavelength radio astronomy was in its infancy. A plethora of molecules was discovered in the interstellar medium, including water, ammonia, formaldehyde, carbon monoxide, ethyl alcohol and a number of molecules not observed terrestrially. It was quickly demonstrated that excitation of the key species implied collisions by neutral hydrogen molecules at high density (by interstellar standards) and at a temperature as low as about 10 K. The problem is that when the gas is cold, molecular hydrogen does not emit an observable quantity of radiation and therefore cannot be directly observed. In-

stead, its presence has to be inferred from trace constituents with abundances at most 10^{-4} that of H_2 . The most widely used tracer is CO because it is the easiest to observe, but estimating H_2 masses from CO observations is like estimating the weight of a dog from the number of hairs on its tail. To make matters worse, CO is an optically thick molecule; the strength of its radio-spectral lines is more closely related to the temperature of the gas cloud than to how many molecules are radiating.

Two significant discoveries in the mid-1970s emphasize the importance of obtaining accurate H_2 masses from the CO observations. The first is the ubiquity of CO in the plane of the Milky Way, implying that a significant fraction, perhaps most, of the interstellar gas is H_2 and not atomic hydrogen (H I), which previously had been thought to be the reservoir of the star-forming gas. The second is that essentially all present-day star formation takes place in giant molecular clouds, which have masses of about 10^5 solar masses and are composed almost entirely of H_2 (but detected by radio observations of CO). Therefore, the question of star formation on a galactic scale was reduced to the question of how giant molecular clouds form. Do they, for example, condense from the atomic gas, or do they form from the collisional accretion of small molecular clouds without the mediation of an atomic phase? Important clues were to be found in the relative amounts of H_2 and H I. (Ultimately the answers can be applied to studies of other spiral galaxies. The evolution of the disk of a galaxy is essentially its star-formation history, which in turn is largely the history of the giant molecular clouds.)

The results of the first CO surveys of the galactic plane implied that between the galactic centre and the position of the Sun, which is about half-way out, the interstellar gas is overwhelmingly molecular. One study suggested it contained ten times as much H_2 as H I, which in turn implied that atomic gas is largely irrelevant to the formation of giant molecular clouds, but these conclusions were quickly challenged by several groups who argued that the molecular phases are about equal in abundance inside the solar distance. In that case,

100 years ago

THE way in which a bird builds its nest, seemingly without instruction, has been brought forward as a proof of blind instinct governing it in its task. A remarkable instance, however, of a changed mode of nest-building has been brought to my notice by Mr W. Burton. His brother took to New Zealand some young chaffinches. Some of the birds have built a nest, and the structure shows very little of that neatness of fabrication for which the bird is noted in England. The cup of the nest is small, loosely put together, and the walls of the structure are prolonged for about eighteen inches, and hang loosely down the side of the supporting branch. Clearly these New Zealand chaffinches were at a loss for design when fabricating their nest.

From *Nature* 31 533, 9 April 1885.

instabilities in the atomic component are likely to play an important role in the formation of the giant molecular clouds.

The different estimations have been largely the result of using different ratios to convert CO line strengths to H_2 column densities, and also to the importance of temperature and heavy-element abundance variations in the Milky Way. New light has been shed on this subject by gamma-ray astronomy. It was established in the 1970s that diffuse high-energy gamma-rays (> 50 MeV) come primarily from the interaction of cosmic rays with the interstellar gas. If the gamma-ray emissivity of the gas can be determined, the number of hydrogen nuclei in a column along a given line of sight can be obtained from measurements of the gamma-ray intensity. The contribution to this intensity from atomic hydrogen is known from 21-cm line observations; the remaining contribution must therefore be from molecular hydrogen (corrections due to helium and other trace constituents are straightforward). In principle, this approach is the most accurate way to determine the H_2 from CO, because the gamma-rays directly trace the dominant component of the gas — hydrogen — independently of excitation and abundance variations in the CO.

Interpretation of the gamma-ray results has not been without its problems, however. Initially, the results from the US SAS-2 satellite and the European COS-B

satellite were not in agreement — a large and difficult to determine instrumental background for COS-B seems to have been the culprit — and other technical problems with the data analysis have only recently been solved. Nevertheless, the quantity of molecular hydrogen in the Milky Way reported by the Durham group on page 511 of this issue is still half that derived by the Caravane collaboration, which launched COS-B. The groups agree, however, that the gamma-ray data unequivocally imply that the H_2 does not dominate the H I inside the solar distance.

More H_2 would simply produce more gamma rays than are observed. The work of the Durham group implies that nowhere in the Milky Way does the H_2 surface density dominate that of the H I. In the published work of the Caravane collaboration, H_2 may dominate weakly within a small range of distance from the galactic centre. The residual difference between the two groups probably results from systematic uncertainties that increase the formal errors assigned to the data by Bhat *et al.* A new sophisticated analysis by the Caravane collaboration using the most extensive available data sets is forthcoming and should shed additional light on this nagging astrophysical controversy. □

Leo Blitz is in the Department of Physics and Astronomy, University of Maryland, College Park, Madison 20742, USA.

Evolution

Interpretations of mass extinction

from Michael J. Benton

IT IS becoming clear that the history of life has been punctuated by a series of events during which, typically, an apparently random selection of organisms died out. These mass extinction events can be seen to have 're-set' the evolutionary clock: they wiped out whole families or orders of plants and animals, irrespective of the darwinian fitness of the individual organisms or of the species involved, and created opportunities for the survivors to radiate out into the 'empty' ecospace that was left. The recent suggestion that mass extinctions might be cyclical in occurrence is important because it raises the possibility of identifying one kind of explanation for all mass extinctions through time, whether or not the events have a regular periodicity and whether the ultimate causes are terrestrial or extraterrestrial. These ideas have important implications for evolutionary biology and for historical geology.

The most important recent impetus to the discussion of the causes of mass extinctions has been the suggestion of Raup and Sepkoski¹, from their analysis of the fossil record of marine animals, of a regular periodicity of about 26 Myr to mass extinctions over the past 250 Myr.

Hallam² has pointed out how uncertainties in dating of mass extinctions, especially before the late Cretaceous, weaken the claim of regular periodicity, but the suggestion remains a very important one and is having a considerable influence on the ways in which we may interpret a cyclical pattern of mass extinctions.

Four interpretations of the cyclicity are possible and worth considering. The first two are deterministic and the other two stochastic viewpoints: (1) each extinction event was triggered by the same external impulse, acting directly; (2) the cyclical pattern of extinction events correlates with some physical phenomenon that follows the same cycles (for example, fluctuations in temperature or sea level), and these are controlled by a third external variable; (3) the events are caused by a variety of factors that relate to the internal dynamics of the biological system; (4) the apparent cyclicity is an artefact of the patchiness of the fossil record.

The first interpretation is favoured by Raup and Sepkoski and their followers, who suggest an extraterrestrial cause for the periodic mass extinctions. Three main schools of thought have sprung up. The

first is that a hypothetical companion star of the Sun (called Nemesis) disturbs the Oort cloud and sends a shower of comets hurtling towards Earth every 30 Myr^{3,4}. The second is that the whole of the Solar System oscillates through the galactic plane with a semi-period of 30 Myr and periodically scatters Oort cloud comets into the terrestrial zone⁵⁻⁷. The third idea is that the unknown Planet X, located beyond the orbit of Pluto, follows an eccentric path which causes it to disturb a comet belt beyond Neptune and precipitate a shower of comets, some of which would eventually reach the Earth, every 28 Myr⁸. These astronomical theories are still highly speculative and controversial (see *News and Views* 7 March p.17).

The second interpretation, that cycles of mass extinction correlate with cycles in terrestrial phenomena, has been the generally held view for some time. Fischer suggested that the Earth has passed through cycles of different orders of magnitude^{9,10}, the largest of which each lasted about 300 Myr and began with rapid convection in the mantle. This led to the break-up of Pangaea-like supercontinents, elevated sea levels and marine transgressions, intense vulcanism, release of CO₂ to the atmosphere and the development of a greenhouse climate, resulting in worldwide warm conditions and warm seas. Such episodes occurred from late Cambrian to late Devonian and early Jurassic to late Eocene.

The second phase of each 300-Myr supercycle was characterized by low mantle convection, continental accretion, low sea level and regression, reduced atmospheric CO₂ and the development of an icehouse state. This gave high latitudinal climatic gradients, cold dry polar regions and cold oxygenated ocean waters. These conditions characterized the latest Precambrian to early Cambrian, the late Palaeozoic to Triassic and the latter half of the Cenozoic, in which we still live. The crossovers from greenhouse to icehouse correspond to mass extinctions in the late Cambrian (500 Myr), the late Devonian (355 Myr), the late Triassic (192 Myr) and the late Eocene (40 Myr). These crossovers are about 150-Myr apart, giving a span of 300 Myr for the full cycle. This model does not explain the terminal Permian and the terminal Cretaceous extinction events.

Fischer also described two classes of shorter-range cycles in tectonic activity and world climate^{9,11}, one set occurring every 30 Myr and the other consisting of smaller climatic cycles in the range 20,000–400,000 years, which are induced by orbital perturbations. The 30-Myr cycles are reflected in temperature fluctuations, variations in carbon isotope ratios and in the fossil record of planktonic and nektonic taxa.

Newell^{12,13} and Hallam^{14,15} have argued that sea-level changes are the most important prime cause of mass extinctions, thus laying the main emphasis on one segment of the Fischer theory. Sea-level changes have exhibited a cyclicity on several scales

through the Phanerozoic and many of the well-known mass extinctions correlate well with worldwide episodes of regression. In general terms this usually coincides with a reduction in the area of diversity of continental shelf habitats, and in some cases (for example, the end-Plinianian, end-Cenomanian and possibly the end-Triassic, late Devonian and end-Ordovician events) with periods of oceanic anoxia, as recorded by extensive black shale deposits.

The other kind of terrestrial explanation, which modifies part of Fischer's thesis and relegates sea-level changes to a minor role, emphasizes changes in temperature. This has been championed recently by Stanley¹⁶⁻¹⁸. He notes evidence of global episodes of cooling that coincide with mass extinctions in the late Eocene and the Pliocene and hints at evidence for temperature changes linked to earlier mass extinction events (it becomes harder to pin down temperature changes the further back in time one goes). Stanley supports this thesis with the fact that mass extinctions have frequently been concentrated in the tropics, which became a refrigerated trap into which high-altitude forms migrated and were then wiped out. He also argues that the long periods of time that characterize most mass extinctions accord more with a gradual change in temperature than with a catastrophic change brought about by an extraterrestrial agent.

If, however, cyclicities in geological processes should turn out to have regular periodicities, an extraterrestrial cause will have to be considered. Has the time come with the discovery by Rampino and Stothers¹⁹ of two dominant periodicities, of about 33 and 260 Myr, in a time-series analysis of various geological phenomena (sea-level fluctuations, sea-floor spreading discontinuities, tectonic activity, geomagnetic reversals)? These correspond to the two major orders of cyclicity identified by Fischer, Rampino and Stothers suggest they were caused by periodic comet impacts. The issue at stake is clearly the evidence for regular periodicity in geological processes, which has been questioned (see *News and Views*, 7 March p.17).

There are important philosophical consequences if we accept some or all of these aspects of regular cyclicity in mantle, plate-tectonic, eustatic, atmospheric and climatic processes and their concomitant effects on the history of life. Periodicities of these kinds would form a part of Charles Lyell's interpretation of Uniformitarianism which has generally been abandoned²⁰. Lyell believed that there were constant changes on the Earth, but that these cycled endlessly and with no direction. He would not accept the idea of progressionism in the history of life, believing instead that any life form could exist at any time if the conditions were right: thus he expected to find Silurian mammals, or dinosaurs, at some future time. This anti-progressionist stance was ignored or abandoned by most other scientists during Lyell's lifetime, and the

idea of uniformity of conditions has also not found favour. Now it must be reconsidered if we accept that most major processes on the Earth have followed, and continue to follow, long-term cyclical patterns.

The third interpretation of cyclical mass extinctions — that they arise from the internal dynamics of the biological system — has recently received support from Kitchell and Pena²¹. Attempting to fit different time-series models to the extinction curve of Raup and Sepkoski, they find that the best statistical fit is a stochastic time-series model that shows 'pseudocycles' every 31 Myr. This model takes account of the observed spacing and the relative magnitude of each peak of mass extinction. Kitchell and Pena conclude that each extinction event should be investigated on its merits; it may be incorrect to seek an overall controlling pattern or an ultimate cause that explains all mass extinctions.

The fourth explanation, which is that the apparent regular pattern is an artefact, has also received support. Re-analysing Raup and Sepkoski's data, Hoffman and Ghild find that they correspond just as well to a neutral model, in which origination and extinction rates vary independently and randomly through time, as they do to the deterministic periodic model²². Episodes of high origination and high extinction rates usually correspond with highly fossiliferous geological units. The apparent periodicity of 26 Myr in mass extinctions could be an artefact of the average length of stratigraphic stages, the basic time units used in the calculations.

We are left with all four explanations for mass extinction cycles. Arguments for and against each view will continue. But more detailed analyses of the fossil data, and refinements in dating events in the geological past, will be necessary before we can decide which view is correct. □

1. Raup, D.M. & Sepkoski, J.J., Jr. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801 (1984).
2. Hallam, A. *Nature News and Views* **308**, 686 (1984).
3. Whitmire, D.P. & Jackson, A.A. *Nature* **308**, 713 (1984).
4. Davis, M., Hut, P. & Muller, R.A. *Nature* **308**, 715 (1984).
5. Rampino, M.R. & Stothers, R.B. *Nature* **308**, 709 (1984).
6. Schwartz, R.D. & James, P.B. *Nature* **308**, 712 (1984).
7. Clube, S.V.M. & Napier, W.M. *Nature* **311**, 635 (1984).
8. Whitmire, D.P. & Matese, J.J. *Nature* **313**, 36 (1985).
9. Fischer, A.G. in *Biotic Crises in Ecological and Evolutionary Time* (ed. Nitecki, M.H.) 103 (Academic, New York, 1981).
10. Fischer, A.G. in *Catastrophes and Earth History* (eds Berggren, W.A. & Van Couvering, J.A.) 129 (Princeton Univ., Princeton, N.J., 1984).
11. Fischer, A.G. & Arthur, M.A. *Spec. Publ. Soc. econ. Paleontol. Miner.* **25**, 18 (1977).
12. Newell, N.D. *Spec. Pap. geol. Soc. Am.* **89**, 63 (1967).
13. Newell, N.D. in *Catastrophes and Earth History* (eds Berggren, W.A. & Van Couvering, J.A.) 115 (Princeton Univ., Princeton, N.J., 1984).
14. Hallam, A. *Nature* **269**, 769 (1977).
15. Hallam, A. *Rev. Earth planet. Sci.* **12**, 205 (1984).
16. Stanley, S.M. *Geology* **12**, 205 (1984).
17. Stanley, S.M. *Scient. Am.* **250**, 46 (1984).
18. Stanley, S.M. in *Mass Extinctions* (ed. Nitecki, M.H.) (Univ. Chicago Press, 1984).
19. Rampino, M.R. & Stothers, R.B. *Science* **226**, 1427 (1984).
20. Gould, S.J. in *Catastrophes and Earth History* (eds Berggren, W.A. & Van Couvering, J.A.) 9 (Princeton Univ., Princeton, N.J., 1984).
21. Kitchell, J.A. & Pena, D. *Science* **226**, 689 (1984).
22. Hoffman, A. & Ghild, J. *Geol. Mag.* **122**, 1 (1985).

Michael J. Benton is in the Department of Geology, The Queens University of Belfast, Belfast BT7 1NN, Northern Ireland, UK.

Anthropology

Improved timing of hominoid evolution with a DNA clock

from Peter Andrews

OF THE many contributions in the past year towards the taxonomic applications of protein, amino-acid and DNA analyses, it is worth singling out the approach of DNA-DNA hybridization, which has been claimed by C.G. Sibley and J.E. Ahlquist (*J. molec. Evol.* 20, 2; 1984) to provide new insights into the evolution of the hominoid primates. How well does this claim stand up to examination?

Molecular approaches to taxonomic studies were initiated by Morris Goodman, who compared the equivalent protein in different species by immunological techniques. This was followed by Sarich's calculations of protein 'distances' from microcomplement fixation data. Others have calculated allelic differences measured by electrophoresis. What these methods have in common is that they produce overall distance measures from the study of proteins, and if they are compared with morphological analyses they are seen to be akin to the numerical taxonomy and multivariate analyses that were performed in the 1950s and 1960s.

Two claims are made for these protein analyses. One is that genetic similarity between taxa can be identified by minimizing the genetic distance between them, and from this the relationships and evolutionary branching patterns of the taxa can be inferred. The other is that the rates of change of the proteins are sufficiently regular for the genetic distances to be applied as a clock. As to the first claim, while it seems likely that the inferred branching patterns are usually correct, it is not possible to test for correctness. This is because there is no way of knowing what properties of the proteins are contributing to the distances measured — the same problem encountered with morphological distances in numerical taxonomy and multivariate analyses for very much the same reasons. The necessary data to overcome this fundamental objection only becomes available when proteins are sequenced (for instance, Goodman *et al. Nature* 303, 546; 1983).

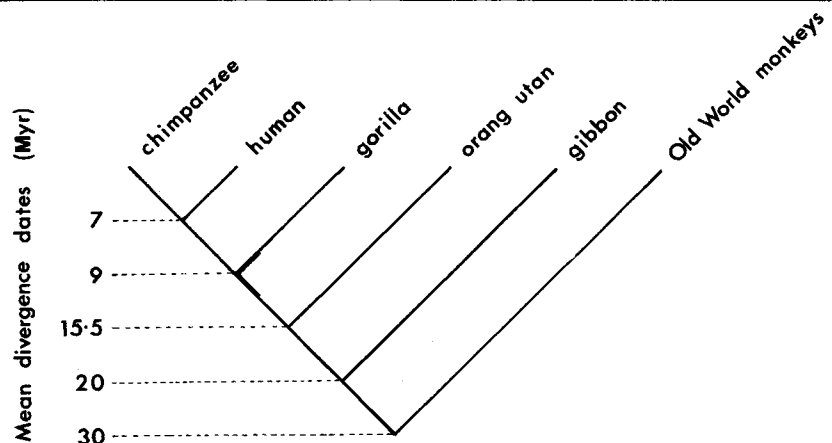
There are also problems with the use of genetic distances as a clock. The clock hypothesis assumes correctness of branching patterns, and goes on to estimate relative times between branching points based on distances observed. The relative times are translated into absolute times by calibration, usually from the fossil record. Taking the example of the Hominoidea, the zoological group to which man and the apes belong, the branching pattern of the various species of hominoid is established according to the protein data, relative times of divergence between man, chimpanzee,

gorilla, orang utan and gibbon are calculated on the basis of the distances between the species and absolute times of divergence are calculated by calibrating one of the relative times with a 'known' divergence date from the fossil record (usually the separation of hominoids from the Old World monkeys).

Two problems affect the accuracy of the protein clock. First, the fossil calibration date may be wrong because it is based on the earliest appearance in the fossil record of a species belonging to a particular lineage. It gives only a minimum date for the origin of the lineage, and the 'earliest appearance' is something that is always subject to change from new evidence. Second, the rates of change of the proteins being analysed may change along or between lineages; Goodman and others have shown that, particularly for the Hominoidea, the rate of change in a number of

similarities in DNA between hominoid species are likely to be the result of shared mammalian or primate inheritance, and only a few will be unique to the hominoids. Therefore, although DNA-DNA hybridization is based on the whole genome, only a very small part of the genome is relevant to hominoid relationships. How to distinguish this part from the rest of the DNA molecule is not stated by Sibley and Ahlquist, but if it is not distinguished how is it possible to test the claims made for it?

This question is not so intractable as it might seem. First of all, it is justly claimed that when the whole genome is tested the numbers of changes and homologies are so great that discrepancies are minimized for more distantly related taxa. But this might not apply to fine resolution. For example, DNA hybridization is claimed to link man more closely with the chimpanzee than the gorilla — a relationship that is not resolved by any of the protein distance methods. In fact, the distances given (with standard deviations) are $1.8-1.9 \pm 0.2$ for man/chimpanzee, 2.4 ± 0.2 for man/gorilla and $2.1-2.3 \pm 0.2$ for chimpanzee/gorilla. This degree of resolution is not impressive, given the size of the standard deviations and that only a small proportion of the genome contains relevant homologies, and I would not



proteins is not constant. The protein clock must therefore be said both to be uncertainly calibrated and to keep irregular time.

Much of what has been said so far applies equally to DNA-DNA hybridization. In this technique, fragments of disassociated single-stranded radioactively tagged DNA from one species are re-associated with single strands from other species; the greater the base-pair similarity between the strands of DNA, the greater the degree of reassociation between them and the greater the force needed subsequently to disassociate them again. Measurement of this force serves as a new measure of overall genetic similarity between the species tested. This similarity is reasonably assumed to be homologous (although inevitably some will be due to homoplasy) but no distinction is made between homologies that are shared uniquely and those that are present in some distant common ancestor. But almost all of the

consider the question of man's closest living relative to be solved by these data.

The major limitation of the hybridization data is overcome by comparing DNA sequences instead, because both the number and kind of differences between species is unambiguous when sequences are compared. Thus, although the hybridization results seem merely to be confirmed by sequences of a short section of the mitochondrial DNA of the hominoids (Brown, W.M. *et al. J. molec. Evol.* 18, 225; 1984 and Hasegawa, M. *et al. Proc. Japan Acad. Sci.* 60, 95; 1984), the sequence data impart a level of confidence that the branching order, shown in the figure, is correct.

Having accepted the branching order, we can return to the DNA hybridization results and consider the value of the DNA clock that was derived from them. Sibley and Ahlquist find a much greater regularity of DNA change than has been demonstrated for protein change, for which there is

evidence of slowing down or speeding up in some cases. To calibrate the clock from the fossil record, Sibley and Ahlquist used two possible dates — 13 and 16 million years — for the divergence of the orang utan. The correct date is probably somewhere in between so that, by taking a range, the other branching points shown on the figure can be estimated.

The conclusion seems to be that DNA hybridization provides a useful insight into evolutionary events, at least for the homi-

noids. The resolution of branching patterns is perhaps not quite as great as has been claimed and, since it is inherently untestable, needs to be tested by sequencing data. More sequencing data will be needed to confirm the identification of man's nearest living relative. The DNA clock seems to keep very good time, in contrast to the protein clock, and gives a divergence time between man and the African apes of between 7 and 9 million years ago. In this connection, the recent description of a new

hominoid from the Samburu Hills of Kenya (Ishida H. *et al. Af. Study Monog.* 2, 73; 1984) is of particular interest. It is placed at about 9 million years ago (the heavy black line on the figure) and is said to be either an immediate ancestor of the African apes and man or an African ape with some similarity to the gorilla. □

Peter Andrews is at the British Museum (Natural History), Cromwell Road, London SW7 5BD, UK.

In my note "Opportunity Lost in 1865?" (*Nature* 306, 313; 1984) about the election of the Chair of Experimental Philosophy at Oxford in the 1860s, I remarked that in 1871 Helmholtz was asked by Sir William Thomson (later Lord Kelvin) whether he would accept the newly founded Cavendish Chair in Cambridge. Since this fact is not widely known, it seems justified to reprint in full Thompson's letter to Helmholtz which is among the Helmholtz papers in The Central Archives of the Akademie der Wissenschaften der Deutschen Demokratischen Republik in Berlin.

The combined university/college emoluments suggested to Helmholtz came to £780 to £800, only slightly more than the £700 which Oxford would have offered to Helmholtz six years earlier if Max Müller had not vetoed the invitation on account of the inadequacy of the proposed salary.

Helmholtz, who at the time of the Cambridge approach was still Professor of Physiology in Heidelberg, declined the offer since in June 1870 he accepted the Chair of Physics at the University of Berlin and was due to take up his appointment in the spring of 1871. There seems to be no record of Helmholtz's answer to Thomson. There is nothing in the Kelvin collection at the Cambridge University Library, but there is a letter from Helmholtz's wife to her mother dated 9 February 1871 (quoted in *Anna von Helmholtz, Ein Lebensbild in Briefen* Verlag für Kulturpolitik, Berlin, 1929) in which she wrote that "Hermann has just received an enquiry from Sir William Thomson whether he would be inclined to accept a Professorship in Experimental Physics in Cambridge, under very attractive conditions — however it is to be Berlin" ("Allein nun ist doch Berlin die Lösung"). Then, in a letter dated 13 February 1871 the chairman of the Appointments Committee wrote to Maxwell asking him to stand as a candidate and expressing the hope that his election would be secured. It seems that these matters were dealt with most expeditiously a hundred years ago and that the postal services were gratifyingly swift.

I thank Dr Christa Kirsten, Director of The Central Archives of the Akademie der Wissenschaften der DDR, for the permission to reproduce this letter and Mr J. Deakin, Secretary of the Cavendish Laboratory, Cambridge for the information about the invitation to Maxwell.

N.Kurti, University of Oxford

Helmholtz's choice

London
Jan 28, 1871

My dear Helmholtz
I have been asked
by Stokes, and by the Master
and Tutor of my College
at Cambridge, to write
to you asking if you could
be induced to accept a
new professorship of
Experimental Physics
to be established there.
It is much desired to
create in Cambridge
a school of experimental

My dear Helmholtz

I have been asked by Stokes, and by the Master and Tutor of my College at Cambridge, to write to you asking if you could be induced to accept a new professorship of Experimental Physics to be established there. It is much desired to create in Cambridge a school of experimental science, not merely by a system of lectures with experimental illustrations but by a physical laboratory in which students under direction of the professor and his assistant or assistants, would perform experiments, and the professor would have all facilities attainable, for making experimental investigations.

The Duke of Devonshire has already given £6000 for building the laboratory and making a commencement of providing it with instruments. I think it may be confidently expected that funds will not be wanting for carrying out this part of the place well. The University proposes to give £500 per annum to the professor, whose income will also be increased somewhat by fees from students. But if you were accepting the professorship it is I may say quite certain that you would also be appointed at least to a fellowship in St Peter's College which would bring an additional income of £250 or £300.

Thus I believe the income to be relied on would be not less than £800. But I think it probable that a praelector-fellowship of Trinity College would be offered to you, although on this point I speak merely from my own judgement. The duties of the praelector-fellowship, a new institution in Trinity College (of which one in physiology has been given to Dr Michael Foster and it

was intended there should be also one in experimental physics) would be fulfilled by giving facilities to members of Trinity College for attending the professorial lectures: that is to say if the university professor is appointed to the praelectorship. Otherwise the praelector would lecture on physics in Trinity College simply. The income of the praelectorship would be about £600 so that the professor if appointed to the praelectorship would have £1100 per annum.

I know that the question of amounts of income is very far from the first in respect to the inducements which might possibly cause you to think of accepting such a proposal: and I am quite prepared to learn that you are "absolute for" Berlin. Still it is a question that must be weighed before you could decide to accept, and therefore I have told you what I can regarding it.

If you could at all entertain the idea of coming to this country, I think that the situation at Cambridge would present many advantages. It would certainly be a most interesting field, in respect to our profession, as the desire for physical science is growing stronger and stronger in the University, and the force of public opinion is steadily advancing in support of it, and to stimulate it where stimulus is needed.

The duties of lecturing would occupy only 20 weeks out of the year, and all the remainder of the time would be available for experimental or mathematical investigation, and scientific writing. Will you give me a line at the earliest, to say if you think you could be induced to accept or if on the contrary you decide at once against it. I need not say that it would be a great gratification and advantage to English scientific men to have you among us instead of merely having very rare opportunities of seeing you, and that I myself would consider the difference of distances from Glasgow to Cambridge and Berlin a great gain.

I write necessarily in haste, and can only add that, irrespectively of this question, I wished to write to you asking for news of yourself and of your family, and particularly regarding how you have been affected by this terrible war. I hope very much that you have not now cause for anxiety in respect to your son or other relations.

I am in London on account of a Committee invited to advise the Admiralty on Designs for Ships of War, especially in respect to stability, in consequence of the loss of the "Captain" which was upset with 500 men in the Bay of Biscay last September. When you write, address The College, Glasgow, as I am in London only for a few days each fortnight.

Believe me

Yours always truly

William Thomson

Yellow rain and the bee

SIR — Criticism of the bee faeces theory by Rosen *et al.*¹ is misleading and inaccurate. Their assignment of the burden of proof to US government sceptics is misplaced.

The authors assert that "heavy" invasion of substrate is a prerequisite for high toxin levels and that they never observed any "fluffy" mycelium in their samples. In fact, the relationship between mycelial mass and toxin production by *Fusarium* spp. is extremely variable, ranging well over three orders of magnitude. Further, the authors neglect to mention that they did not microscopically examine their samples. While at the University of Wisconsin, I examined bee faeces samples from Thailand. No mycelial growth was obvious but *Fusarium* spp. were present at counts as high as 1,000 colony forming units per milligram. At least one of these strains is capable of producing greater than 1,000 p.p.m. zearalenone under laboratory conditions. It is curious that the authors choose to subject their samples to highly sophisticated GC-MS analysis without also determining the mycoflora present by dilution plating and/or microscopy.

Rosen *et al.* claim that Type A and Type B trichothecenes are "never (except in one case)" found together. This "lone exception" is purported to be a corn sample collected in France. As at least one of the authors should know, this is incorrect².

In their final paragraph, Rosen *et al.* state "We are amazed that the proponents of this theory have not yet presented one iota of evidence that trichothecenes can be produced naturally . . .". When one considers the resources available to the US government's laboratories and their intense efforts to identify additional positive samples, it is truly "amazing" that none have been found. More importantly, it is astonishing that the authors imply that the burden of proof should lay with the sceptics. Meselson and others have merely advanced an alternative explanation that is consistent with all physical evidence. In contrast, the US State Department has termed its evidence as "unequivocal" and the so-called "smoking gun".

This "conclusive" evidence has been the basis of serious allegations of Soviet treaty violations and has directly affected such issues as the binary weapons programme and disarmament negotiations. Under those circumstances, who has the burden of proof?

DANIEL CULLEN

180 Kimball Way,
South San Francisco,
California 94080, USA

1. Rosen, J.D., Cohen, H., Mirocha, C.J. & Schiefer, H.B. *Nature* 313, 271 (1985).
2. Mirocha, C.J., Schauerhamer, B., Christensen, C.M. & Koddedahli, T. *Appl. Envir. Microbiol.* 38, 557-558, (1979).

Fivefold symmetry in boron and borides

SIR — A recent News and Views article presented an interesting discussion of fivefold crystal symmetries¹, in particular relating to the recent experiments on Mn-Al alloys². I would like to draw attention to the fact that the fivefold symmetry is also a characteristic feature of the crystalline boron (beta-rhombohedral modification) as well as some borides (such as AlB_2). The crystal lattice of β -boron is formed by the repetitive pattern of icosahedrons with the unit cell having 105 non-equivalent atoms. This is by itself rather peculiar for an elemental material. This fact, strikingly little known outside the boron specialists, leads to a number of non-trivial consequences. One example is the possibility of hopping conductivity in an intrinsic (and not doped) semiconductor³. Therefore, the naturally occurring fivefold symmetry has another area of manifestation, besides the Mn-Al alloys mentioned.

A.A. BEREZIN

Department of Engineering Physics,
McMaster University, Hamilton,
Ontario, Canada L8S 4M1

1. Maddox, J. *Nature* 313, 263 (1985).
2. Shechtman, D. *et al.* 53, 1951 (1984).
3. Berezin, A.A. *J. chem. Phys.* 80, 1241 (1984).

King Midas and the Red Queen

SIR — Benton's commentary¹ on recent tests of the Red Queen hypothesis draws attention to the difficulty of distinguishing changes in the biotic environment from those in the abiotic one, over geological time. The hypothesis demands that evolution is driven by an antagonistic biota. However, mutualistic interactions among species cause the biotic environment to improve rather than to deteriorate², and the need for continual change just 'to keep in the same place' is lost. Thus, an alternative way of testing the hypothesis is to compare the rates of evolution of taxa in antagonistic environments with those found in mutualistic ones, specifically in associations where one species lives inside the cells of another. We suggest that the homeostatic intracellular environment experienced by mutualistic endosymbionts is buffered from antagonists while, at the same time, subject to much the same major abiotic changes (such as climatic perturbations) as the environment of their hosts.

Nature has been kind enough to provide many such evolutionary experiments, including mutualistic symbioses as diverse as algae within coelenterates (for example, corals), and bacteria within plants (for example leguminous and non-leguminous nodules, see Table 1 of ref. 2). A striking feature of these and other mutualistic associations is that the symbionts which live intracellularly for a prolonged period are represented by a much smaller number

of taxa than their external partners². We argue that this supports the view that biotic interactions do play an important part in evolutionary processes^{2,3}.

For those inclined to regal metaphors, we would contrast the Red Queen of deteriorating, antagonistic environments, with King Midas's golden touch of species in improving, mutualistic environments, although the outcome is evidently more stable for mutualistic species than it was for the king himself³.

R. LAW

Department of Biology,
University of York,
York YO1 5DD, UK

D.H. LEWIS

Department of Botany,
University of Sheffield,
Sheffield S10 2TN, UK

1. Benton, M.J. *Nature* 313, 734 (1985).
2. Law, R. & Lewis, D.H. *Biol. J. Linn. Soc.* 20, 249 (1983).
3. Law, R. in *The Biology of Mutualism: Ecology and Evolution* (ed. Boucher, D.H.) (Croom-Helm, London, in the press).

Biological role of IgG hinge region

SIR — We feel compelled to point out that an article correlating IgG antibody flexibility and complement fixation by the classical pathway¹, while an important piece of work, is a regrettable example of authors failing to acknowledge the intellectual and experimental contributions of earlier workers. Surely the findings of Oi *et al.* must be considered a quantitative refinement of principles laid out originally by us in a series of papers over the past decade (for example refs 2-6). The following⁵, for example, is unambiguous:

(These studies . . . highlight the crucial role played by the hinge region and its constituent disulfides in allowing IgG to perform its biological effector functions. An essential feature of this role is to limit segmental flexibility; yet this flexibility seems to be important for optimal antigen binding. Thus, the hinge serves to balance the antigen-dependent and antigen-independent flexibility requirements of the molecule.

It is surprising that the editorial review practices of *Nature* can be so accommodating to what we believe an imbalanced presentation.

K.J. DORRINGTON
D.E. ISENMAN
M.H. KLEIN
R.H. PAINTER
D.G. ROMANS

Department of Biochemistry,
University of Toronto,
Toronto, Canada M5S 1A8

1. Oi, V.T. *et al.* *Nature* 307, 136-140 (1984).
2. Isenman, D.E., Dorrington, K.J. & Painter, R.H. *J. Immun.* 114, 1726-1729 (1975).
3. Romans, D.G., Tilley, C.A., Crookston, M.C., Falk, R.E. & Dorrington, K.J. *Proc. natn. Acad. Sci.* 74, 2531-2535 (1977).
4. Romans, D.G., Tilley, C.A. & Dorrington, K.J. *Molec. Immun.* 16, 859-879 (1979).
5. Klein, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 78, 524-528 (1981).
6. Dorrington, K.J. & Klein, M.H. *Molec. Immun.* 19, 1215-1221 (1982).

The World Ocean Circulation Experiment

J. D. Woods

Institut für Meereskunde an der Universität Kiel, Düsterbrookweg 20, D 2300 Kiel 1, FRG

The international oceanographic community is preparing for an ambitious five-year experiment. Starting in 1990, it will provide the first comprehensive global survey of physical properties of the oceans. The resulting data set will be used to test computer models of the ocean circulation needed for predicting decadal climate change. The results will also benefit research in marine chemistry, biology and geology. This review looks at the climatological problems that motivated the project and discusses new observing techniques, from space and in situ, which make it timely to embark on this major step in physical exploration of the ocean.

THE World Ocean Circulation Experiment (WOCE) is being planned to survey the global distributions of ocean variables with a view to greatly improving estimates of the circulation of heat, water and chemicals around the world ocean, and their exchange with the atmosphere. Such experiments become necessary when the existing database is inadequate for a particular purpose; a new data set stimulates new thinking and reveals deficiencies in old ideas and models. A recent example was the Global Weather Experiment (GWE) in 1979, which has had a major impact on improving models used in weather forecasting¹. The aim of WOCE is to collect a data set that will do the same for oceanic circulation. It will provide the first comprehensive global perspective of the ocean as an element in the planetary climate system. The experiment will open a new chapter in the history of ocean exploration, as the *Meteor* Atlantic expedition did in 1925–27, the *Discovery* Southern Ocean expeditions in the 1930s, the International Geophysical Year (IGY) Atlantic expedition in 1957 and the International Indian Ocean expedition in 1966 (ref. 2).

This article opens with a brief survey of the climatological problems that motivate the WOCE, then considers the present state of ocean circulation modelling, and surveys the data available to test them. The urgent need for a new global data set becomes evident. Finally, the aims, options, strategy and design of the experiment intended to meet that need are discussed.

Decadal climate change

There are many potential applications of a detailed understanding of the global circulation of the oceans, but the driving force for attempting a WOCE is the recognition that predicting decadal climate change will depend on accurate calculation of changes in the large-scale flow of heat, freshwater and chemicals in the oceans. The global pattern of climate on the continents is strongly influenced by those fluxes³. Estimates based on the

difference between the radiation budget at the top of the atmosphere and the fluxes carried by the atmosphere⁴ have revealed that the meridional heat fluxes in the ocean and the atmosphere both lie in the range 0.1–1 PW (10^{15} W), with the ocean contributing more in the tropics and the atmosphere more at higher latitudes (Fig. 1). Attempts are being made to extend the calculations with a view to discovering how much the flow in each ocean basin contributes to this zonally averaged heat flux^{5–7}. This indirect approach has been stimulated by concern that calculations based on maps of surface energy fluxes may be unreliable because of errors in the bulk formulae used to calculate the fluxes from ship meteorological observations and uneven distribution of and errors in the observations. Global maps of heat and freshwater circulation based on the best available estimates are characterized by large flows between the basins (Fig. 2)⁸; the mean energy and water supply to the atmosphere in the Atlantic depend in part on solar heating and rainfall in the Pacific. Attempts have been made to calculate the fluxes more directly from hydrographic data at a few sites⁹ (Fig. 3); the method gives values in the same range (0.1–1 PW) but differing significantly in detail. The great width of the Pacific Ocean presents special problems, and there is disagreement about the direction of the fluxes in some regions.

The bulk formula method is suspect because it gives unrealistic budgets for the Mediterranean and Red Sea¹⁰; the direct method can be used reliably only at locations where the flux calculation is insensitive to the unknown barotropic component of the geostrophic current; the residual method is suspect because calculations of the atmospheric flux divergences give unrealistic values over the continents. The discrepancies between the three methods are quite modest by comparison with those in other aspects of planetary physics, and might be thought of as presenting no urgent challenge. But we are not concerned here with the energy balance of some distant planet. We are discussing the energy and water budgets which govern the climate of Earth¹¹. The impact of climate change on mankind provides the motivation for predicting future changes in climate. Climate prediction provides challenging specifications for the accuracies to which we must measure ocean fluxes.

Consider the North Atlantic. The incoming heat flux (1.2 PW) is lost to the atmosphere at a basin-average rate of 50 W m^{-2} . The flow is unevenly distributed, but errors in the bulk method prevent us from mapping the distribution more accurately than $\pm 50 \text{ W m}^{-2}$. The global patterns of surface heat and water fluxes are not constant; they exhibit a broad spectrum of variability. There has long been speculation¹² that such oceanic changes cause some of the observed variations in climate over the continents, but because of the deplorable inaccuracy of the methods used to measure the surface fluxes, there is usually no hope of deducing the changes from past records. Exceptions are sites such as the Eastern Equatorial Pacific, where interannual variation exceeds 50 W m^{-2} (ref. 13).

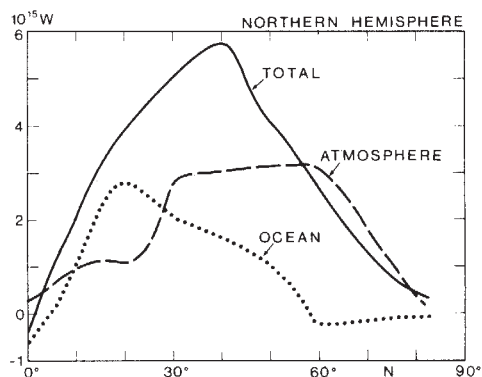


Fig. 1 The meridional heat transport carried by the ocean and the atmosphere⁴.

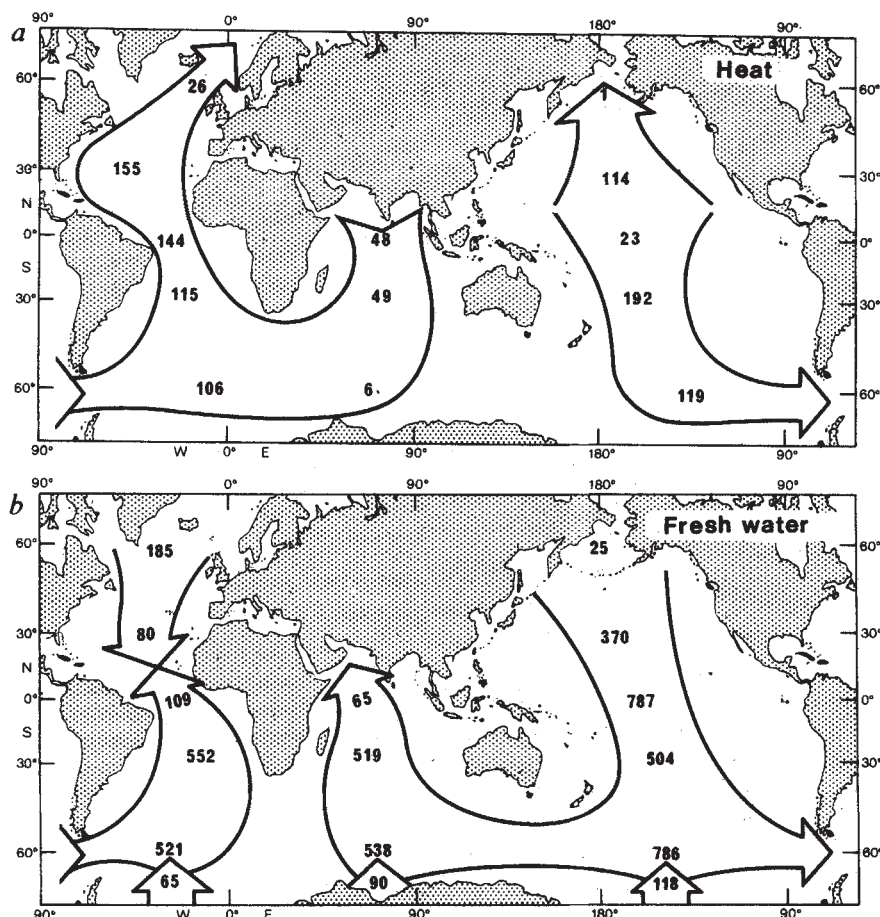


Fig. 2 Global circulations of heat (a, 0.01 PW) and fresh water (b, $10^3 \text{ tonnes s}^{-1}$) based on surface fluxes estimated from ship weather observations⁸.

Even quite small changes in the surface fluxes averaged over large areas can have a large impact on climate if they persist for decades. They can be provoked by changes in the solar constant or in the atmospheric concentrations of dust or gases. The palaeoclimatological record derived from polar ice cores contains evidence of such changes¹⁴. The forcing might involve quite a small deviation in the surface radiation flux; for example, the natural doubling of atmospheric P_{CO_2} during the past 15,000 years changed the surface infrared flux by about 10 W m^{-2} (ref. 15). Sensitivity tests with atmospheric climate models have shown that important consequences can result from such forcing¹⁶.

When we consider the possibility of forecasting decadal climate change, we become concerned with the processes inside the ocean (circulation and mixing) that provide the link between changes in surface forcing and the resulting changes in energy, water and gas fluxes to the atmosphere. Here we concentrate

on the energy flux, although for some purposes the change in gas flux (especially CO_2) is also important. The scientific problem is to calculate the thermal response of the ocean to an externally induced change in surface radiation balance. The forcing changes the heat content of the ocean, leading to a change in the surface temperature and consequently in the rate of oceanic cooling to the atmosphere. Given a step change in the radiation input, the oceanic temperature will relax to a new equilibrium value, for which the globally averaged surface energy flux budget is again balanced. The new balance will involve a different partition between the components that make up the total cooling to the atmosphere; long-wave radiation, conduction (sensible heat flux) and evaporation (latent heat flux). The climate is sensitive to that partition. An increase of solar constant or of P_{CO_2} will lead to increased evaporation balanced by greater precipitation, some of which will fall on the continents.

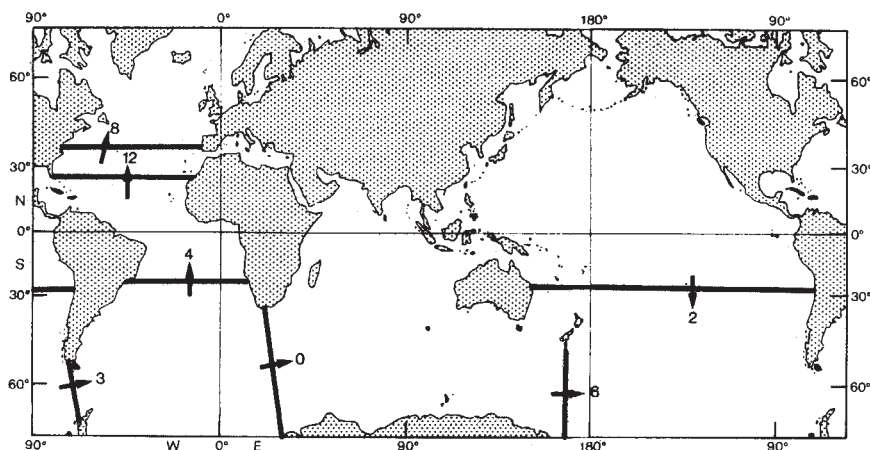


Fig. 3 Direct measurements of heat flow (0.1 PW) in the ocean based on hydrographic sections and other oceanographic measurements⁹.

Decadal changes of ocean heat content are not confined to the surface mixed layer (mean depth 70 m), but are ventilated by deep winter convection and by Ekman pumping to depths of the order of 1 km (ref. 17). The involvement of a greater mass of water prolongs the relaxation time, giving ocean currents scope to redistribute the change of heat content over great distances. As a result, the equilibrium change in surface cooling and evaporation may be distributed quite differently from their mean patterns and from the distribution of surface forcing. Furthermore, the transient change may have different characteristics in different parts of the world, and the annual mean response may involve different seasonal changes at different locations.

To summarize, the global patterns of seasonal and annual climate—in particular rainfall—will change over decades in response to radiative forcing of the ocean by natural variations of solar constant and atmospheric dust and gas concentrations. These lead to changes in ocean heat content. There is no prospect of predicting decadal climate changes until it is possible accurately to model the action of oceanic circulation and mixing in redistributing the decadal heat content anomalies, as an essential prerequisite to calculating the changes in surface fluxes of energy, water and gases to the atmosphere¹⁸.

A computer model designed to predict decadal climate change must, therefore, include the ocean circulation and mixing, as well as the atmospheric circulation and physics of clouds, radiation and precipitation. No such comprehensive model exists at present, although several groups are working towards that end¹⁹. The atmospheric component of such a model is well advanced, but the oceanic component lags behind. Ultimately the two components must be integrated because of feedback between the atmospheric climate and the oceanic circulation. It will not be satisfactory to treat the ocean currents as a stationary feature of a changing climate, but that refinement is of secondary importance compared with the urgent task of creating a model that can at least simulate the circulation of the ocean forced by a steady atmospheric climate. The availability of such a model would revolutionize our ability to diagnose the role of the ocean in our present climate, and would act as a springboard for developing the coupled atmosphere-ocean models needed to predict decadal changes of climate.

The strategy for oceanographers contributing to the problem of predicting decadal climate change is to develop a model that accurately describes the circulation of heat, fresh water and selected dissolved chemicals (notably those involved in the carbon cycle), using a description of the present seasonal cycle of atmospheric forcing; to test this model against a description of the circulations derived from observations; and progressively to improve the model until it predicts the fluxes of energy, water and gases (notably CO₂) into the atmosphere with errors significantly smaller than the observed decadal fluctuations.

Modelling the ocean circulation

The aim of WOCE is to collect a data set that will test and stimulate the development of models of ocean circulation needed for predicting decadal climate change. The class of model to be considered here is sometimes called 'prognostic', but it is not yet used in a forecasting mode, as that would require coupling to an atmospheric model. It is used in a diagnostic mode, being designed to calculate the seasonally varying distributions of currents, temperature, salinity and chemical concentrations around the world ocean, given the surface fluxes of momentum and vorticity, radiation, heat, water and gases as functions of season and geographical distribution. One of the triumphs of oceanography in the past decade has been the successful development of such a model at Princeton²⁰. Starting with a uniform motionless ocean, the model has been integrated for the equivalent of 1,000 years with prescribed surface boundary conditions to give realistic patterns of surface geostrophic flow (Fig. 4).

The output of such a model integration can be used as a synthetic data base for the world ocean, with the desirable

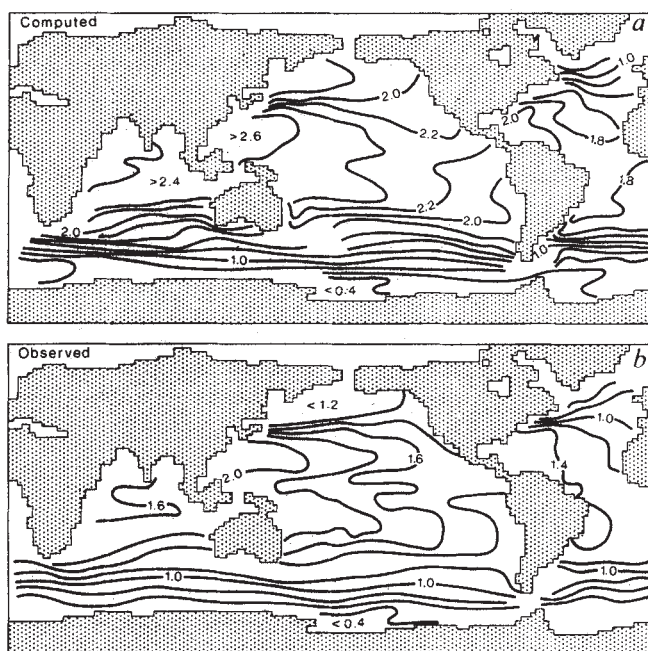


Fig. 4 The distribution of surface elevation of the ocean relative to the geoid (a) from the Princeton model (b) deduced from the observed density distribution using an assumption that the velocity is zero at 2,000 m (ref. 20). The surface geostrophic flow goes along the isolines, with a speed given by the horizontal gradient of surface elevation (that is, hydrostatic pressure).

attribute that there are none of the data-sparse regions that plague analysis of real data. For example, it is possible to calculate the fluxes of heat and water through trans-ocean sections²¹. The results agree remarkably with those derived from analysis of observations (notably across the Atlantic at 23°N), giving us confidence in the model. Although encouraging, are the predictions sufficiently accurate to meet the needs of climate forecasting? And if not, can we identify and correct the sources of error in the way that meteorologists have systematically improved models used in weather forecasting? We do not have an adequate data base to answer such questions fully; that is why oceanographers need a WOCE, just as meteorologists need the GWE. We can, of course, spot obvious differences between the distributions of temperature and salinity predicted by the model and those printed in oceanographic atlases²², but we cannot say whether differences are significant for prediction of decadal climate. However, most oceanographers are concerned about the difference in structure of the permanent thermocline (Fig. 5), since that is likely to be symptomatic of an error in the rate of ventilation of heat from the surface into the interior of the ocean, a process of central importance for decadal climate change^{23,24}.

One reason that the world ocean model gives a false thermocline structure may be that, to run it on the Princeton computer (one of the most powerful in the world), the spectrum of motion has had to be truncated at 250 km. The energetic eddies which constitute the oceanic equivalent of weather systems were not resolved; their effects on the velocity, temperature, salinity and chemical distributions had to be parameterized in the model equations by Boussinesq eddy transport coefficients (Fig. 6). Furthermore, the large truncation scale meant that rather large values had to be used for those coefficients; it was not possible to simulate the observed regional variation of eddy kinetic energy by prescribing regional variation of the coefficients, although sensitivity tests showed that the model results were affected by arbitrary (homogeneous) variations in their magnitude²⁰. As meteorologists had learned with the much broader atmospheric weather systems, the problem could be largely eliminated by using a computer powerful enough to resolve the eddies. The arrival of a new generation of supercomputers in

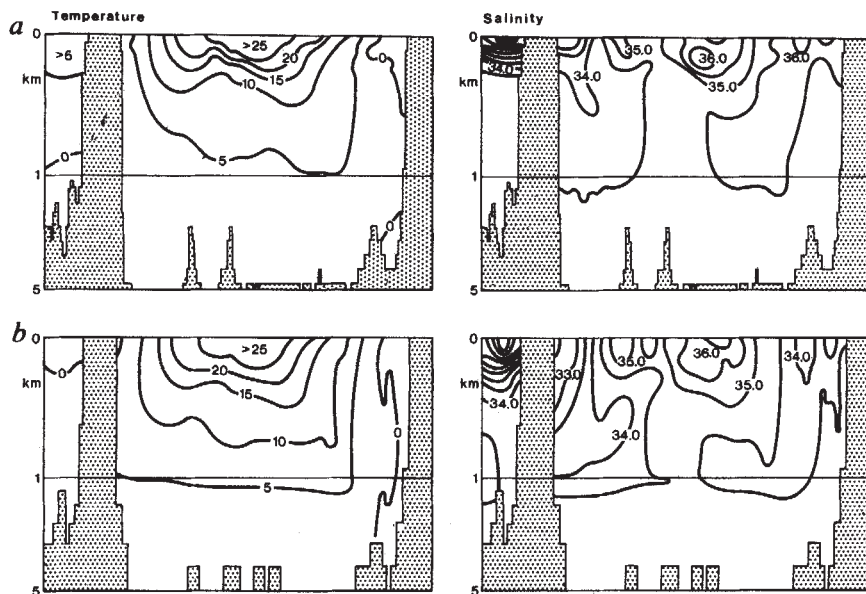


Fig. 5 The distribution of temperature and salinity in a vertical meridional section through the Pacific Ocean: *a*, from the Princeton model; and *b*, from hydrographic data²⁰. The similarities in *a* and *b* are encouraging, but the difference in the depth of the 10 °C isotherm indicates that the model does not describe upper ocean ventilation adequately for decadal climate prediction.

the mid-1970s made it possible to test that hypothesis for a small ocean basin. Holland²⁵ demonstrated that the eddies influence the mean circulation and showed that the mean circulation is affected by bottom topography and the density distribution through the action of the eddies. Several modellers are exploring more sophisticated parameterizations designed to permit progress without resolving the eddies (J. S. A. Green, personal communication), but it is generally believed that they will have to be resolved in future climate models. The present generations of computers makes that impossible except in limited regions, or idealized ocean basins. But in designing WOCE we are looking forward to models that will be in operation at the end of the century. By then there will be two more generations of computers, with at least a factor of 100 in power, making global eddy-resolving models a practical proposition.

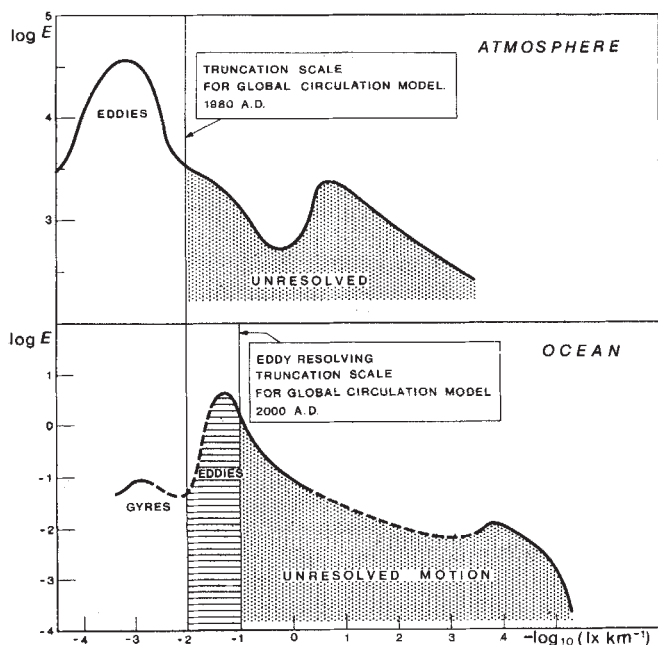
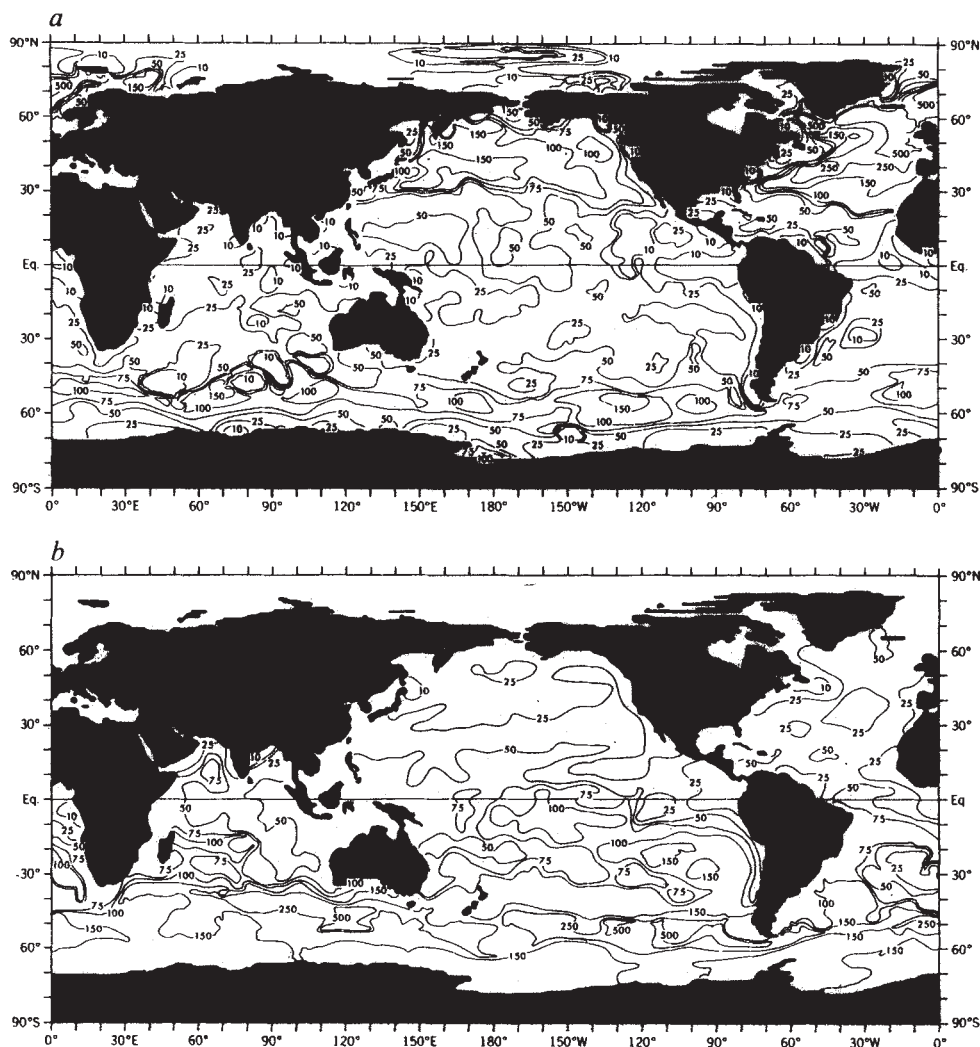


Fig. 6 Typical spectra of motion in the ocean and atmosphere showing the energy peaks associated with eddies. The eddy energy must be resolved in models of the global circulation. Existing computers permit a truncation scale of 100 km, which resolves eddies in the atmosphere, but not in the ocean. The continuing increase in computer power will permit progressively smaller truncation scales until, by the end of the century, it should become possible to resolve the ocean eddies.

Another source of error is the parameterization of vertical mixing in the statically stable interior of the ocean. In the world ocean model, this is based on a vertical eddy diffusivity for scalars and a vertical eddy viscosity for momentum. Research into small-scale transport processes has cast doubt on the efficacy of such parameterization. While many details remain to be settled, there are several processes with different transport characteristics: scalar mixing is effected by billow turbulence²⁶, double diffusive convection²⁷ and cabelling²⁸; momentum transport is effected by internal waves²⁹ and quasi-geostrophic eddies³⁰. Analysis of scalar distributions measured during large-scale geochemical surveys have emphasized the need to distinguish between vertical and diapycnic mixing³¹. Numerical techniques have been proposed to avoid this source of error by coordinate transformation³² or, more drastically, by reformulating the circulation models in terms of isopycnic rather than isobaric surfaces³³. Such techniques will not only improve the prediction of scalar diffusion by unresolved motions, but they also simplify the treatment of potential vorticity conservation, which is the theoretical linchpin of eddy and gyre scale dynamics³⁴.

Climate models are particularly sensitive to the parameterization of boundary layer processes, which control the deep ventilation of anomalies entering the ocean at the surface, and, conversely, the mixed layer temperature and therefore the fluxes of energy, water and gases into the atmosphere³⁶. The standard practice in ocean circulation models is to treat the boundary layer physics one-dimensionally³⁷. That is useful for simulating the response of sea-surface temperature and salinity to the seasonal variation of surface energy and water fluxes, but it does not take account of advection of heat and water by ocean currents, the essence of global climate¹⁷. Nor is it possible with such boundary layer parameterization to predict the depth of the mixed layer in winter, when surface water mixes deepest (Fig. 7). As a result, it is not possible to calculate the regional and temporal variation of annual water mass transformation in the boundary layer, a crucial mechanism for decadal climate change. Recent work by the Princeton group has revealed the sensitivity of deep-water ventilation to the depth of winter mixings^{23,24}. It was shown that simulation of the observed distribution of tritium (from nuclear bomb tests) in the North Atlantic was greatly improved by artificially making the mixed layer in winter have the regional distribution deduced from observations. The need for accurate parameterization of boundary layer processes will become increasingly important when climate models are based on freely coupled ocean-atmosphere circulations, for which the present expedient of prescribing the seasonal variation of mixed layer temperature, salinity and depth will no longer be appropriate.

Fig. 7 The climatological mean distribution of the depth of the oceanic mixed layer at the end of winter, as deduced from density profiles⁴⁴, has been used as a parameter in models designed to simulate the penetration of anomalies from the atmosphere into the interior of the ocean^{23,24}.
a, March; b, September.



Future models

Increasing computer power will improve the resolution of ocean circulation models and include more elaborate parameterizations of unresolved processes. Faced with such opportunities, the modeller has to ask how much emphasis to give to processes competing for each increment in computer power. His decisions are influenced by three factors: (1) theoretical ideas illustrated by process models; (2) the availability of data to run the model with new features; and (3) the availability of data to test the resulting model predictions. It is also important that the modeller has a specific goal in mind.

The ultimate goal for modelling in WOCE is decadal climate prediction, which emphasizes those aspects of ocean circulation that influence oceanic feedback to the atmosphere over time-scales of 10 or 100 years, and stresses the coupling of the global circulations of the ocean and atmosphere. The theoretical ideas that will most strongly influence WOCE model development will be those concerned with the oceanic boundary layer, the ventilation and circulation of the thermocline (the warm water sphere), and the fluxes between basins which give the ocean its global integrity. The problem facing the modeller is to assimilate the many advances made in recent decades in our understanding of processes that influence these aspects of the large-scale circulation, and to decide how much weight to give to each as the model evolves. A powerful theoretical tool for making such decisions is provided by Ertel's concept of potential vorticity conservation³⁴. The full implications of the concept are still being investigated, but it has already been shown that most aspects of the dynamical state of the ocean can be determined if one knows the distribution of potential vorticity along isopyc-

nic surfaces and the density distribution along the boundary³⁵. By focusing on the distribution of potential vorticity along isopycnic surfaces, it becomes possible to assess the relative importance (for large-scale circulation) of a multiplicity of transient motions, including the quasi-geostrophic eddies and fronts, convection in the boundary layer (due to surface cooling) and in the interior (due to double diffusion), turbulence in the mixed layer and the interior, and so on. It seems likely that potential vorticity will be the dynamical touchstone of WOCE, just as geostrophy was for the great oceanographic expeditions of previous generations.

The data needed to run ocean circulation models include: (1) the topography of the sea floor as lower boundary condition; (2) the distributions of velocity, temperature, salinity and selected chemicals as initial conditions; and (3) the global patterns of seasonally varying fluxes of momentum, energy, moisture and gases at the ocean surface interface as upper boundary conditions. The bathymetry is already known to the required resolution of 30 km in some regions^{38,39}, but remains undersampled in most of the ocean. It is significant that 70 unknown seamounts were discovered by analysis of the Seasat altimeter data^{40,41}; geophysicists predict many more. The accurate mapping of passes through which deep water passes between ocean basins presents a continuing problem.

In a recent census of the world ocean waters, Worthington⁴² assessed the accumulated hydrographic data base available in the World Data Centres. There are large gaps in the Southern Hemisphere that have to be spanned by interpolation. The uncertainty in the mean heat content and total mass of salt is important even when the model is integrated from a

'homogeneous, motionless' initial state. The alternative of initializing the model, as meteorologists do in weather forecasting, with the best available description of the present day distributions is more challenging⁴³. That requires a knowledge of the global distributions. The first systematic world analysis of temperature, salinity and oxygen distributions (with a resolution of $1^\circ \times 1^\circ \times 33$ layers) has recently been published⁴⁴. The potential of surface pressure maps from altimetry for initialization of ocean circulation models is being investigated⁴⁵. Looking further ahead, acoustic tomography⁴⁶ may provide the synoptic data required for model initialization.

The upper boundary condition for ocean circulation models used in decadal climate prediction will come from the atmospheric general circulation models (GCMs) coupled to them. At present, there are design faults in atmospheric GCMs which lead to large errors in the surface fluxes they predict¹¹. Attention is now being given to identifying and eliminating the causes of those errors. Meanwhile, the ocean modeller must proceed in the uncoupled mode, using observed surface fluxes as the upper boundary condition. At present, they are calculated from ship meteorological observations using formulae which have been calibrated against measurements made during air-sea interaction experiments such as JASIN 1979 (ref. 47). The unacceptably large uncertainties in the surface fluxes so determined have been widely discussed³. They represent a major source of error in predictions of ocean circulation models. There is a very real risk of rubbish in: rubbish out. Nevertheless, the global analysis of seasonally varying surface fluxes derived from ship observations represents the state of the art. Precipitation is notoriously difficult to estimate at sea and published distributions have to be treated with caution^{48,49}. Some modellers, faced with the uncertainty in maps of surface fluxes of energy and water, have used the better observed distributions of surface temperature and salinity. That has permitted useful progress with many aspects of circulation modelling, but it is one stage further removed from the goal of coupled atmosphere-ocean climate models. The early attempts at developing coupled models have been based on a simpler parameterization of the surface fluxes based on the newtonian cooling law¹⁹.

Summarizing, the existing descriptions of bathymetry, hydrographic distributions and surface fluxes needed to run ocean circulation models already pose difficulties and it is becoming increasingly urgent to improve them so as to avoid the risk that ocean circulation models with greatly improved resolution and physics fail to realize their potential because of inadequacies in the data used to run them. Improvements are needed in accuracy, spatial and temporal resolution and in global coverage.

Testing the models

There are two strategies for using observations to test models. The first focuses on products closely related to what the customer wants, in our case decadal climate forecasts. The second focuses on internal properties of the system being modelled; does the circulation look right? The weather forecaster can follow the first approach. The absence of adequate time series from the past, and the need to make predictions before another century has passed, forces the oceanographer to concentrate on the latter. In designing WOCE, we must choose to observe those aspects of the ocean circulation that offer the best prospect of testing the plausibility of models that will be working at the end of this century.

The requirement is for global models that treat the circulation around the world ocean in its totality. But the present state of oceanographic knowledge is uneven, and even the major advance expected as the result of WOCE cannot realistically be expected to raise our level of knowledge of all regions to the highest level. Special emphasis will be placed during WOCE on filling the most glaring gaps in our data base, but the resources available are limited and will be deployed in a way that offers the best prospect of testing future models, not simply on a regular grid. The pragmatic criterion for deciding what constitutes useful testing of a new model depends on pushing back

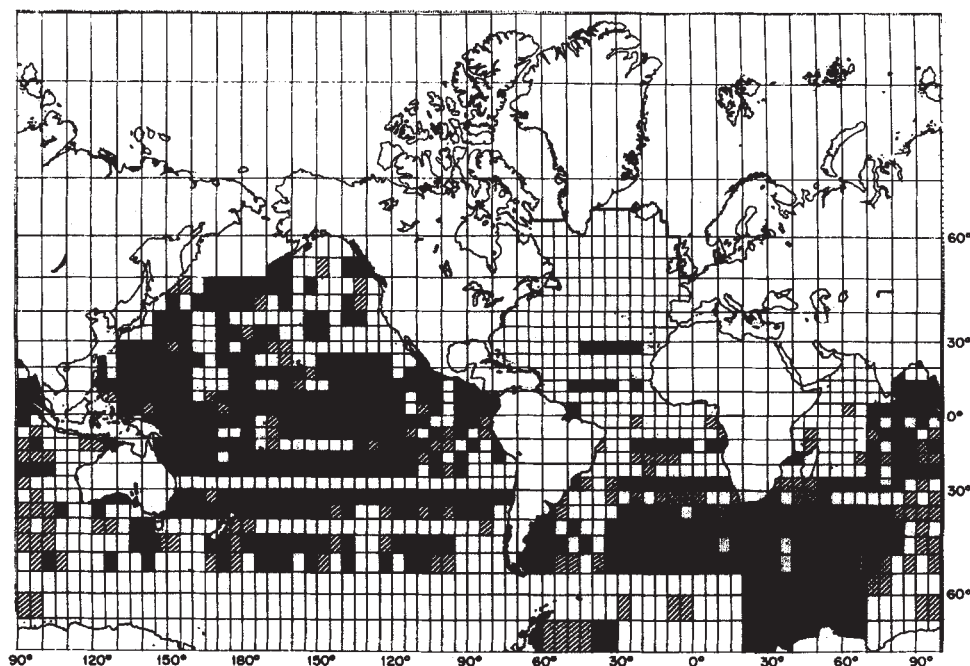
the frontiers of our state of knowledge. The detailed specification varies considerably from region to region. In some places, such as the North Atlantic, our knowledge of the circulation is relatively advanced and we can proceed to specify errors in measurements of, for example, the heat flux. Elsewhere, we are still relatively ignorant of the general character of the circulation; for example, we have neither documentation nor theory concerning the variability of the Antarctic circumpolar current, and we simply do not know how it connects with the circulations in the Pacific, Atlantic and Indian Oceans to the north. In such cases, WOCE must be designed to reveal the unknown; comparison with model predictions will have to be more qualitative than quantitative. The major discoveries of WOCE are expected to come from the Southern Hemisphere, which has been poorly sampled in the past, for obvious logistic reasons.

Broadly, the procedure will be to extract products from the WOCE data set that can be compared with equivalent products generated by the prognostic models. At the simplest level, it is useful to compare the distributions of temperature, salinity and dissolved chemicals (see Fig. 5). The same goes for the patterns of mean and variance of the surface pressure distribution measured by satellite altimetry. The velocity field presents greater difficulties because it must be calculated from a diagnostic model supplied with scalar distributions. Classical attempts at diagnostic modelling (reviewed in ref. 50) involved arbitrary assumptions. Recently, there have been significant advances in the application of inverse theory to the design of diagnostic models that extract products from oceanographic data, with explicit statements about the errors in the data and the tolerance to which dynamical constraints apply⁵¹. Once the velocity field has been diagnosed, it is possible, in principle, to compute the fluxes of heat, water and chemicals, and to establish the relative contributions of the different classes of motion to those fluxes (that is, the Ekman transport in the mixed layer, the wind-driven gyre circulations, the meridional circulation driven by deep convection at high latitude, and the eddies). By comparing the observed distributions of dissolved chemicals with those computed using a model that incorporates the WOCE empirical velocity field, it will be possible to assess the stationarity of the circulation on the ventilation timescales of the different regions, and so resolve the present uncertainty as to the origin of water in the deep Pacific; is it a fossil of an earlier climate or in equilibrium with our present climate? Inverse theory warns us that it will not be easy to determine velocities and fluxes accurately. In general, even with the great increase in data we expect from WOCE, the system will be under-determined, and model testing will be more effective if we concentrate on collecting data that will permit us to calculate a few products accurately. Compared with the richness of detail meteorologists have following their global experiment, the number of such products expected from WOCE may seem discouragingly small, but they will provide powerful constraints on the models at the global level.

On a smaller scale, it is necessary to test the parameterizations of motions that cannot be resolved even in the models running on twenty-first century computers. Here the comparison will be between observed large-scale indicators of those motions, for example, the annual maximum depth of the mixed layer, which is crucial in parameterization of the ventilation of properties acquired in the boundary layer into the interior of the ocean¹⁷. Comparison of isopycnic potential vorticity distributions will provide one of the crucial tests of the model parameterizations of the effects of internal mixing on large-scale circulation. It is emphasized, however, that WOCE will not incorporate any detailed process studies related to particular classes of mixing: the experiment will deploy all available resources on the largest scales of motion.

Summarizing, the existing database tests models by revealing errors in scalar distributions, but provides very little information about the velocity field and the large-scale fluxes. New data are needed to extend the existing tests globally and to make it possible to test the velocity field, fluxes and internal dynamics of the models.

Fig. 8 The global distribution of data available for running ocean circulation models is biased towards the Northern Hemisphere, shipping lanes, fisheries zones, and against the winter months when extra-tropical air-sea interaction is strongest. The distribution of high-quality hydrographic data⁴² is shown as an example. Black areas contain no suitable data; hatched areas contain data only in shallow locations.



Aims of WOCE

The discussion of model testing has revealed serious deficiencies in the existing database, which can be grouped into three categories.

First, inadequate coverage of the world ocean. The Southern Hemisphere data coverage is sparse at all seasons; everywhere there is poor coverage in winter compared with summer (Fig. 8). Merchant ship observations, which are crucial for calculating surface fluxes needed to run the models, are concentrated on the shipping lanes and biased against rough weather, when much of the air-sea interaction may be occurring. There is an urgent need for a homogeneous global data base covering all seasons.

Second, although the large oceanographic data archives available at the World Data Centres still contain untapped information, the individual measurements were often collected with other objectives in mind (such as to solve regional circulation problems) and are not well suited to testing global circulation models designed for climate prediction. There are many hydrocast profiles, but few eddy-revolving, top-bottom, trans-ocean sections.

Third, the accuracy of many measurements does not meet the specification generated by analysis of climate prediction requirements, because measurements were made originally for other purposes, for example the surface fluxes derived from ship meteorological observations. More frustrating, because it arises from sloppy practice, is the poor salinity accuracy in deep profiles from some laboratories.

The aim of the WOCE is to remedy those deficiencies by new measurements, designed specifically to meet the needs of global circulation models for climate prediction. It will take advantage of the remarkable advances that have been made in theoretical understanding of ocean circulation, in instrumentation and in design concepts since the last great international expeditions⁵². Above all, the aim will be to collect a data set that will permit a global perspective in studies of ocean circulation.

Timing of WOCE

Those are very ambitious aims and the question is, is it timely to embark on such a major project only five years hence? The answer is that four significant advances in the past decade have made it possible to take this logical next step in the physical exploration of the oceans: (1) exploration of the transient, quasi-geostrophic eddies, (2) the first computer model of the world ocean circulation, (3) new techniques for exploring the ocean from space and *in situ*, (4) recognition that improved

knowledge of ocean circulation was needed before the climate impact of CO₂ pollution could be predicted. These four developments will now be discussed briefly in the context of the WOCE. **Eddies.** The importance of the transient quasi-geostrophic eddies for ocean circulation had not been recognized when the last international expeditions were being planned, even though, with hindsight, there is evidence of the eddies in earlier data. Following the POLYGON, MODE and POLYMODE experiments in the past decade and associated theoretical studies, it is possible to assess the role of the transient eddies in the larger-scale permanent gyres⁵³. It is also possible to design an observing strategy and data analysis techniques that take account of the eddies, whether as a dynamically important signal in their own right or as a noise on the gyres. The form of the spectrum of oceanic motion is known to a wavelength of 10 km, which is probably sufficient for designing WOCE. The recognition that general circulation models will have to resolve the eddies is a major factor in planning the experiment.

Models. The successful completion of the first computer model of the world ocean circulation, with the three-dimensional distribution of temperature and salinity, was an important milestone in oceanography. It provides the springboard for successive improvements that will eventually lead to climate prediction models. Future improvements, including resolution of the eddies, diapycnic rather than vertical mixing, and coupling of circulation and boundary layer physics, were discussed earlier. The implementation of those improvements in global models will depend on the availability of adequate computer power. The spatial resolution problem poses the greatest difficulty; a 100-fold increase in computer power is needed to achieve resolution equivalent to that currently used in weather forecasting models. Happily, there is every prospect of that being achieved by the end of this century. The most powerful machines in the world today, the Cray-XMP and Cyber 205, can perform several hundred million instructions per second; laboratories leading research into eddy-resolving ocean circulation models are already planning to order computers that can perform more than a thousand million instructions per second. Even if some of the more bullish predictions of the rate of increase in computer power prove optimistic, there is no doubt that early in the next century, ocean circulation modellers will need a data set to test global eddy-resolving models. The WOCE cannot be completed before 1995 at the earliest and, judging by the time it has taken to complete the GWE data set, the WOCE data set will barely be ready for use by the end of the century—none too soon for testing models using fifth-generation computers.

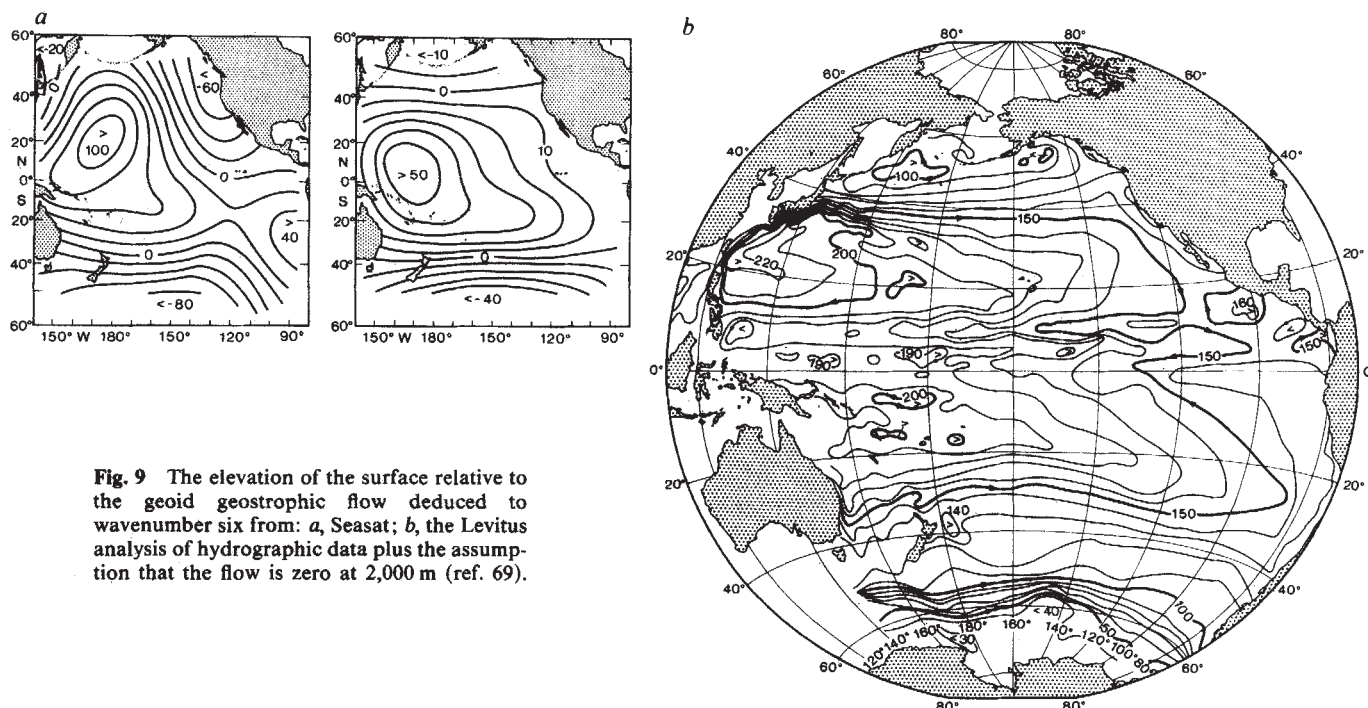


Fig. 9 The elevation of the surface relative to the geoid geostrophic flow deduced from wavenumber six from: *a*, Seasat; *b*, the Levitus analysis of hydrographic data plus the assumption that the flow is zero at 2,000 m (ref. 69).

New tools. The Atlantic expedition during the IGY was based almost entirely on instruments lowered from research ships. The most revolutionary tool was the bathythermograph, which could give a shallow temperature profile without stopping the ship. The IGY data set is the mainstay of contemporary calculations of Atlantic circulation. Research ships will continue to have the central role in ocean exploration for the foreseeable future, and they will provide the core of the WOCE data set. New techniques will improve the efficiency of ship use. The increasing use of expendable instruments (bathythermograph, current profiler) has made it possible to supplement the limited research fleet with ships of opportunity. Systematic surveying of temperature anomalies has been undertaken in this way during North Pacific Experiment⁵⁴, and is now being used in the Tropical Ocean and Global Atmosphere⁵⁵ project.

During the 1970s there was also a rapid development of moorings and drifters for circulation studies. Moorings can now be deployed for up to a year, making it possible to explore variability with periods much longer than the eddy lifetimes⁵⁶. Surface drifters tracked by satellite⁵⁷ and deep drifters tracked acoustically⁵⁸ can be followed for many months. Pressure gauges deployed on the ocean floor for several months have been used to map the tides⁵⁹. The accumulated number of buoy-months has already made it possible to map velocity mean and variance in the North Atlantic⁶⁰. Theoretical studies have helped to clarify the interpretation of the quasi-langrangian tracks of these drifters⁶¹.

Methods of chemical analysis of water samples collected by research ships have advanced rapidly in the past decade. It is possible to measure the concentrations of chemicals present at very low concentrations with far smaller samples than hitherto. And many of the chemicals can now be analysed at sea. Mass spectroscopy has revolutionized the measurement of tritium and other elements previously detected by counting radioactive decay^{62,63}. Those advances have made it possible to undertake extensive surveys of chemical distributions in the world ocean: first GEOSECS (Geochemical Ocean Sections Study)⁶⁴ and more recently TTO (transient tracers in the ocean)⁶⁵. These chemicals supplement the variables (temperature, salinity, oxygen and nutrients) used traditionally to infer the circulation^{52,66}.

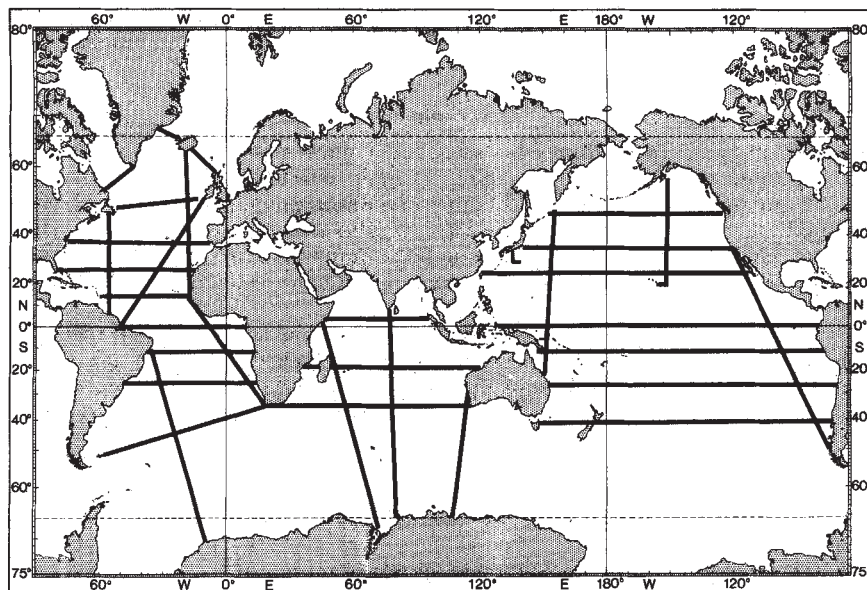
The most spectacular advance in ocean observation has come from the demonstration by Seasat during a 100-day mission in 1979 that satellite radar and passive microwave techniques now offer the accuracy needed for ocean circulation research^{67,68}.

The success of the radar altimeter in measuring the shape of the ocean to a precision better than 10 cm opens the door to global mapping of the surface geostrophic velocity (Fig. 9), thereby allowing oceanographers to dispense with the classical and unwarranted assumption that the flow goes to zero at some arbitrary deep level⁶⁹. The advent of satellite altimetry and plans for future altimetric missions provided the initial impetus for WOCE. The technique offered the prospect of a truly global perspective of the ocean circulation for the first time. The timing of the experiment has from the start been based on the launch date of the next altimetric mission. At present, it seems likely that the European Earth Resources Satellite ERS-1 and the American-French Topex-Poseidon will be launched in 1990 and operated for 3–5 years. The success of other satellite techniques has also influenced the early planning of WOCE. In particular, the radar scatterometer, which promises to provide systematic global measurements of the wind stress on the ocean for the first time⁷⁰. The measurement of sea-surface temperature by infrared radiometer has had a chequered past, but at last seems to be yielding measurements accurate to 0.6 K, with the prospect of improvement to about 0.3 K in the future⁷¹; that begins to be useful for WOCE.

Other techniques showing promise for WOCE include acoustic tomography, which has mapped eddy patterns in an early sea trial⁴⁶, and acoustic Doppler current profiling, which might be deployed widely on merchant ships to map currents in the top few hundred metres of the ocean⁷². Increasing attention is being paid to collecting data from research ships moving at speed (in contrast to the classical technique of profiling on station): undulating profilers derived from the Batfish have been used to collect upper ocean sections several megametres long⁷³. Baker⁷⁴ has reviewed these and other new techniques that lead us to believe that some of the old problems preventing oceanographers from observing the ocean globally are being solved.

CO₂ pollution. Early calculations of the climate response to CO₂ pollution of the atmosphere suggested that equilibrium with twice the present concentration would be achieved when the atmosphere was on average about 2 K warmer than now, with a smaller rise in the tropics but up to 8 K at high latitudes, and an associated change in the pattern of precipitation⁷⁵. There was speculation that the temperature rise might cause the West Antarctic ice sheet to melt, raising sea level by several metres⁷⁶, and that the change in rainfall pattern might have a major impact

Fig. 10 One possible distribution of trans-ocean hydrographic sections for WOCE based on a realistic estimate of ship time likely to be forthcoming from existing resources available to the world oceanographic community (ref. 78). Additional ship resources will be needed to undertake the WOCE hydrographic and tracer programme in the Southern Hemisphere, which is not adequately covered by the sections identified in this map.



on agriculture in some countries⁷⁷. These predictions were sufficiently alarming to create a demand for research into predicting decadal changes of climate (CO_2 concentration being predicted to double within 50 years). Further modelling made it clear that the rate and global pattern of atmospheric temperature rise will be largely determined by sea-surface temperature response to the changing infrared flux from the atmosphere. It was realized that early models based on heating limited to the mixed layer, which gave a response time of a few years, were wrong. It is necessary to take account of the subduction of water from the mixed layer into the interior by Ekman pumping and deep convection in winter¹⁷. That spreads the heat input over a much greater water depth, delaying the attainment of an equilibrium sea-surface temperature rise for decades, and allowing the ocean circulation to distribute the thermal anomaly widely, perhaps even between ocean basins. Unfortunately, the present state of ocean circulation modelling, and especially the treatment of the subduction process, is inadequate to the task of predicting accurately the transient thermal response of the ocean to the changing surface infrared flux due to CO_2 pollution. The early timing of WOCE is, in part, a response to the sense of urgency generated by those concerned with predicting climate response to CO_2 pollution.

Experimental strategy

The primary goal of the experiment is to develop ocean models useful for predicting decadal climate change and to collect the data necessary for testing them. The specific objectives are to determine and understand on a global basis the following aspects of the world ocean circulation and their relation to climate: (1) The large-scale fluxes of heat and fresh water and their divergences within 5 years, and in addition their annual and interannual variability. (2) The inputs and fluxes of potential vorticity. (3) Components of oceanic variability on months to years, megametres to global scale, and the statistics on smaller scales. (4) The rates and nature of formation, ventilation and circulation of water masses that influence the climate system on time scales ranging from 10 to 100 years.

The experimental design strategy being followed for WOCE offers three perspectives on the circulation: the global scale; the links between the separate ocean basins; and inside the basins. Inevitably, there will be overlap between those three perspectives, but the detailed experimental design, and the precise specification for similar measurements, will be different.

The field programme will be intensive, with a limited duration related to the planned mission lives of the next generation of ocean observing satellites, which provide the first opportunity to collect global observations. The expected launch year of those

satellites is 1990, and their planned lifetimes are 3–5 years. The intensive oceanographic programme of WOCE will, therefore, concentrate on the period 1990–95, with the possibility of adjustment according to changes in satellite launch dates. Those measurements in the ocean that need to be made simultaneously with the satellite measurements must be concentrated into that period. Other measurements (for example, some aspects of the deep tracer surveys) can safely start earlier and continue after the 1990–95 period.

The fact that the intensive field programme will extend over several years does not mean that the resulting data set will constitute an average or synoptic description of the world ocean during that period. The satellite data will come closest to the ideal of a time series of synoptic observations. Most of the *in situ* observations will be made once, sequentially as part of a rolling programme extending over 5 or more years. A few measurements will be repeated to check for significant changes, and there will be a small number of time series. There is no single name that describes such a non-synoptic sample, which will inevitably alias most of the changes occurring within the intensive field phase. The word 'snapshot' captures something of the spirit in which the WOCE data set will have to be analysed, provided one bears in mind the photographic problems of camera shake, focal plane distortion, and so on. Despite these problems of interpretation, the WOCE snapshot will be far more coherent and closer to being synoptic than the existing data archive accumulated over several decades. It will, therefore, represent a significant advance in sampling technique, as well as in global coverage. A conscious effort will be made in designing the experiment to minimize the remaining problems of aliasing. Selected measurements will also be continued for a longer period to discover how representative was the WOCE snapshot of the ocean, given its broad spectrum of natural variability.

WOCE design

The development of an experimental design soon becomes a juggling act between the desirable and the possible. This section outlines the principal ingredients of the experiments, as proposed by Wunsch⁷⁸.

(1) The fields of surface forcing as determined by (i) satellite scatterometers (anticipated spacecraft are the European Space Agency's ERS-1 and the US National Remote Ocean Sensing Satellite NROSS) for wind stress; (ii) conventional meteorological analyses for wind stress and heat and moisture fluxes based on improved atmospheric modelling efforts and an upgraded surface observational system using drifters in otherwise uncovered regions.

(2) The field of oceanic surface topography (the oceanic surface pressure boundary condition) determined by (i) satellite altimeters (anticipated spacecraft are ERS-1 and the US NASA Topex Mission); (ii) tide gauge network on coasts and oceanic islands.

(3) Global hydrography from ships using both highest quality conductivity-temperature-pressure profilers plus developing free-fall instruments from 'research ships of opportunity' to provide both a climatological temperature-salinity data base plus continent-to-continent, top-to-bottom sections for use with dynamic method and ancillary observations and models (Fig. 10).

(4) Global measurements from ships (probably combined with hydrographic measurements) of chemical tracers, including tritium, fluorocarbons, oxygen and nutrients. These observations, along with the hydrography, provide the primary observable parameters representing the integrated (that is, time-average) behaviour of the ocean circulation.

(5) Global, but sparse, sampling of selected chemical tracers such as ^{14}C as part of the determination of time-integrated oceanic behaviour.

(6) Global, but sparse, determination of the dominant source-sink terms of the biogeochemical tracers involving transfers to and from the sea floor and continental margins by sediment traps and so on, for purposes of using the tracers.

(7) Direct observations of the flow fields over large areas for long time periods by 'clouds' of neutrally buoyant floats and surface drifters, both communicating directly with satellites.

(8) Measurements of large areal averages for long times, of heat content, velocity and vorticity and potential vorticity through moored acoustic tomography arrays.

(9) Ship of opportunity programmes involving near-surface velocity measurements through acoustic backscatter and direct near-surface density field measurements through expendable bathythermographs.

(10) A few special, regional foci involving ships, conventional moored current meters, floats, temperature recorders, tide gauges and other *in situ* instrumentation.

It is expected that it will take about 5 years to complete this measurement programme. A related programme will monitor critical aspects of the ocean over a longer period to relate that 'snapshot' of the circulation to the broad band spectrum of ocean variability.

Potential for new techniques

The WOCE design outlined above depends heavily on new techniques. How safe is it to base an experiment on new technology, and what is the criterion for rejecting promising but as yet unproven techniques? Experience gained during the GWE is relevant. The core measurements for GWE were the World Weather Watch routine synoptic observations; the equivalent for the WOCE are the research ship hydrographic and chemical surveys. Satellites were crucial for GWE, as they will be for WOCE; if present plans for altimeter and scatterometer satellites were cancelled, the WOCE would have to be postponed. Special observing systems planned for the GWE included neutrally buoyant balloons and buoys drifting on the ocean; the former failed, but the latter were successful and transformed the ability of meteorologists to analyse weather systems in the Southern Hemisphere. Special observing planned for WOCE include acoustic tomography and neutrally drifting buoys that pop up to be interrogated by satellite; both lie at the leading edge of oceanographic technology. Planners of the GWE counted on being able to improve the quality and data returns of well-established routine measurements on land; WOCE needs to do this for the meteorological measurements made by voluntary observing ships.

Looking back on the GWE, it is clear that all the observing systems that were eventually deployed in 1979 had been identified 10 yr earlier. Gaining from that experience, WOCE planning has, therefore, proceeded on the basis that no completely new techniques announced during the 1980s would be ready in

time for WOCE. Most of the techniques that have been listed as needed for WOCE are the subject of intense development efforts. Perhaps the greatest uncertainty concerns the development of a new temperature-salinity profiler that can be used from research vessels not equipped with deep winches. If successful, this new instrument will make it possible to share the enormous task of the WOCE hydrographic programme among a much larger number of research institutes.

Novel techniques will be tested during the WOCE snapshot to discover their potential for cost-effective long-term monitoring of the ocean in the next century.

Conclusion

The World Ocean Circulation Experiment is an undertaking planned to support research in decadal climate variation. It will also provide a unique global perspective of ocean circulation with applications in many other aspects of marine science.

The WOCE is a major component of the World Climate Research Programme (WCRP) sponsored jointly by intergovernmental and non-international governmental organizations. The WCRP is concerned with discovering whether it is possible to predict climate on timescales of several weeks to several decades, and to determine the effect of man's influence on climate. The programme is divided into three streams, concerned respectively with climate prediction in periods of (1) months, (2) years and (3) decades. The critical scientific problem limiting our ability to predict decadal climate change is the inadequate description of and ability to model the circulation of the world ocean. The organizers of the WCRP have therefore proposed the WOCE as the principal activity within stream (3) of the WCRP. They have established the WOCE Scientific Steering Group (SSG) to plan and organize the experiment. The tasks of the SSG are to identify the primary goals and objectives of WOCE, to design a core research programme including modelling and fieldwork, to identify the resources needed to undertake the programme, and to work with international and national bodies, with research institutes and universities and with individual scientists to achieve that core research programme and to encourage the development of associated projects that support the goals of WOCE. J.D.W. is co-chairman of the SSG.

1. GARP Joint Planning Staff *The Global Weather Experiment* (World Meteorological Organization, Geneva, 1981).
2. Schlee, S. *The Edge of an Unfamiliar World: a History of Oceanography* (Dutton, New York, 1973).
3. Woods, J. D. in *The Global Climate* (ed. Houghton, J.) 141-187 (Cambridge University Press, 1984).
4. Oort, A. H. & Vonder Haar, T. H. *J. Phys. Oceanogr.* **6**, 781-800 (1976).
5. Burridge, D. M. in *The Cage Report* (WCP-22, WNO, Geneva, 1982).
6. Campbell, G. C. thesis, Colorado State Univ., Fort Collins (1981).
7. Swinbank, R. *Met. O. 20 Technical Note 11/213* (Meteorological Office, Bracknell 1983).
8. Stommel, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2377-2380 (1980).
9. Bryden, M. L. in *Time Series of Ocean Measurements* (ed. Ellett, D.) (WCP Rep. No. 21, World Meteorological Organization, Geneva, 1982).
10. Bunker, A. F., Charnock, H. & Goldsmith, R. A. *J. mar. Res.* **40** (Suppl.), 73-84 (1984).
11. Dobson, F. E. *et al.* *WCP Rep. Ser. No. 2* (World Meteorological Organization, Geneva, 1981).
12. Namias, J. & Cayan, D. R. *Science* **214**, 869-976 (1981).
13. Niiler, P. P. in *WCRP Pub. Ser. 1*, 473-485 (1982).
14. Dalmas, R. J., Ascencio, J. M. & Legrand, M. *Nature* **284**, 155-157 (1980).
15. Ramathan, V. *J. Atmos. Sci.* **38**, 918-930 (1981).
16. Bryan, K., Komro, F. G., Manabe, S. & Spelman, M. J. *Science* **215**, 56-58 (1982).
17. Woods, J. D. in *Coupled Atmosphere-Ocean Models* (ed. Nihoul, J. C. J.) Ch. 34 (Elsevier, Amsterdam, 1985).
18. Bretherton, F. P. *Prog. Oceanogr.* **11**, 93-129 (1982).
19. Nihoul, J. C. J. (ed.) *Coupled Atmosphere-Ocean Models* (Elsevier, Amsterdam, 1985).
20. Bryan, K. & Lewis, L. J. *J. geophys. Res.* **34**, 2503-2517 (1979).
21. Bryan, K. *Rev. Earth Planet. Sci.* **10**, 15-38 (1982).
22. Stommel, H. & Fieux, M. *Oceanogr. Atlases* (Falmouth, Massachusetts, 1980).
23. Cox, M. D. & Bryan, K. *J. Phys. Oceanogr.* **14**, 674-687 (1984).
24. Sarmiento, G. *J. Phys. Oceanogr.* **13**, 1924-1939 (1983).
25. Holland, W. R., Harrison, D. E. & Semtner, A. J. in *Eddies in Marine Science* (ed. Robinson, A. R.) 379-403 (Springer, Berlin 1983).
26. Woods, J. D. *J. Fluid Mech.* **32**, 791-800 (1968).
27. Schmitt, R. *J. Phys. Oceanogr.* **11**, 1015-1026 (1981).
28. Turner, J. S. in *Evolution of Physical Oceanography* (eds Warren, B. & C. Wunsch) 236-263 (MIT Press, Cambridge, 1981).
29. Oibers, D. *Rev. Geophys. Space Phys.* **21**, 1567-1606 (1983).
30. Rhines, P. & Holland, W. R. *Dyn. Atmos. Oceans* **3**, 289-326 (1979).
31. Rooth, C. & Östlund, G. *Deep-Sea Res.* **19**, 481-492 (1972).
32. Redi, M. H. *J. Phys. Oceanogr.* **12**, 1154-1158 (1982).
33. Bleck, R. & Boudra, D. B. *J. Phys. Oceanogr.* **11**, 755-770 (1981).
34. Holland, W. R., Keffer, T. & Rhines, P. B. *Nature* **308**, 698-705 (1984).
35. Hoskins, B., McIntyre, M. & Robertson, J. *J. R. met. Soc.* (submitted).
36. Woods, J. D. *WCRP Publ. Ser. 1*, 147-179 (1983).
37. Kraus, E. B. (ed.) *Modelling and Prediction of the Upper Layers of the Ocean* (Pergamon, Oxford, 1977).

38. Hydrographer of the Navy *Bathymetry of the Northeast Atlantic* (Taunton, 1975-83).
39. *General Bathymetric Chart of the Oceans* 5th edn (UNESCO, Paris, 1977-82).
40. Dixon, T. H., Narachi, M. & McNutt, M. K. *J. geophys. Res.* **88**, 1563-1571 (1983).
41. White, J. V., Sailor, R. V., Lazarewicz, A. R. & LeSchack, A. R. *J. geophys. Res.* **88**, 1541-1551 (1983).
42. Worthington, L. V. in *Evolution of Physical Oceanography* (eds Warren, B. & Wunsch, C.) 42-69 (MIT Press, Cambridge, 1981).
43. Robinson, A. R. *et al.* *Nature* **309**, 781-783 (1984).
44. Levitus, S. *Climatological Atlas of the World Ocean* (NOAA Prof. Pap. No. 13, U.S. Department of Commerce, Rockville, 1982).
45. Marshall, J. C. *Ocean Modelling* **57**, 7-11 (1984).
46. The Ocean Tomography Group *Nature* **299**, 121-125 (1982).
47. Charnock, H. & Pollard, R. T. *Phil. Trans. R. Soc. A* **308**, 221-449 (1983).
48. Mintz, Y. in *Precipitation Measurements from Space* (eds Atlas, D. & Thiele, O. W.) (NASA Goddard Laboratory, 1981).
49. Baumgartner, A. & Reichel, E. *The World Water Balance* (Elsevier, Amsterdam, 1975).
50. Sarkisyan, A. in *The Sea* Vol. 6 (eds Goldberg, E. D., McCave, I. N., O'Brien, J. J. & Steele, J. H.) 363-458 (Wiley, New York, 1977).
51. Wunsch, C. in *The Global Climate* (ed. Houghton, J. T.) 189 (Cambridge University Press, 1984).
52. Wüst, G. "Meteor" Report VI (1) (English edn by Emery, W. J., 1978) (National Science Foundation, Washington, D.C., 1935).
53. Robinson, A. R. (ed.) *Eddies in Marine Science* (Springer, Berlin, 1983).
54. Wyrtki, K., McLain, D. & Patzert, W. *Variability of the Thermal Structure in the Central Equatorial Pacific Ocean* (Hawaii Institute of Geophysics Rep. HIG-77-1, Honolulu, 1977).
55. Gill, A. E. (ed.) *Toga* (WCP Rep., World Meteorological Organization, Geneva, 1983).
56. Ellett, D. J. (ed.) *Time Series of Ocean measurements* (WCP-21, World Meteorological Organization, Geneva, 1982).
57. McNally, G. J. *J. geophys. Res.* **86**, 8022-8030 (1981).
58. Rossby, H. T., Riser, S. C. & Mariano, A. J. in *Eddies in Marine Science* (ed. Robinson, A.) 66-91 (Springer, Berlin, 1983).
59. Cartwright, D. E. *Proc. 9th int. Symp. Earth Tides*, New York (Springer, Berlin, 1983).
60. Richardson, P. L. *J. geophys. Res.* **88**, 4355-4367 (1983).
61. Haidvogel, D., Robinson, A. R. & Rooth, C. in *Eddies in Marine Science* (ed. Robinson, A.) 481-491 (Springer, Berlin, 1983).
62. Jenkins, W. J. *J. mar. Res. (Suppl.)* **40**, 265-290 (1982).
63. McCartney, M. S. *J. mar. Res. (Suppl.)* **40**, 427-464 (1982).
64. Broecker, W. in *Evolution of Physical Oceanography* (eds Warren, B. & Wunsch, C.) 434-460 (MIT Press, Cambridge, 1981).
65. Roether, W. *WCRP Publ. Ser.* **1**, 459-471 (WMO, Geneva, 1983).
66. Worthington, L. V. *On the North Atlantic Circulation* Ch. 13 (Johns Hopkins University Press, Baltimore, 1976).
67. Kirwan, A. D., Ahrens, T. J. & Born, G. H. *J. geophys. Res.* **88**, 1529-1552 (1983).
68. Woods, J. D. *Phil. Trans. R. Soc. A* **309**, 337-359 (1983).
69. Tai, C. K. & Wunsch, C. *Nature* **301**, 408-410 (1983).
70. Guymer, T. H. *Phil. Trans. R. Soc. A* **309**, 399-414 (1983).
71. Harries, J. E., Llewellyn-Jones, D. T., Minnett, P. J., Saunders, R. W. & Zavody, A. M. *Phil. Trans. R. Soc. A* **309**, 381-395 (1983).
72. Crocker, T. R. *Deep-Sea Res.* **30**, 449-467 (1983).
73. Leach, H., Minnett, P. & Woods, J. D. *Deep-Sea Res.* (in the press).
74. Baker, D. J. in *Evolution of Physical Oceanography* (eds Warren, B. & Wunsch, C.) 396-433 (MIT Press, Cambridge, 1981).
75. Manabe, S., Bryan, K. & Spelman, M. J. *Dyn. Atmos. Ocean* **3**, 393-426 (1979).
76. Wilson, A. T. *Nature* **201**, 147-149 (1964).
77. Kellogg, W. in *Climate Variations and Variability: Facts and Theories* (ed. Berger, A.) 601-621 (Reidel, Dordrecht, 1981).
78. Wunsch, C. in *Proc. Workshop on Global Observation and Understanding of the General Circulation of the Oceans* (U.S. National Academy of Sciences, Washington D.C., 1983).

ARTICLES

Cosmic γ rays and the mass of gas in the Galaxy

Chaman L. Bhat*, Mohamed R. Issa†, Brian P. Houston,
Chris J. Mayer & Arnold W. Wolfendale

Physics Department, University of Durham, Durham DH1 3LE, UK

We make the case that the cosmic-ray particles that produce cosmic γ -rays in the 100-MeV region can penetrate most of the dense clouds of gas that seem to contain much of the mass of gas in the Galaxy. We determine a probable form for cosmic-ray intensity as a function of Galactic position and use the measured γ -ray intensities to derive the surface density of gas as a function of position. Our estimate of $\sim 6 \times 10^8 M_{\odot}$ for the total mass of molecular hydrogen compared with $1 \times 10^9 M_{\odot}$ for atomic hydrogen in the inner Galaxy is a significant change in the contemporary perspective.

THE state of the interstellar medium (ISM) has been investigated in all wavelength regions. We describe here the distribution of gas on the galactic scale from γ -ray studies and put forward other recent supporting evidence.

Information about the magnitude of gas density and its division between the major components, atomic H I and molecular hydrogen, is relevant to several problems in astronomy, one being the manner in which the cosmic-ray intensity varies with position in the Galaxy. The gas distribution must be known to interpret the diffuse γ -ray flux. γ Rays, produced by the interaction of cosmic rays with gas in the ISM, are an ideal tracer because they are not absorbed by the gas.

The distribution of H I is known reasonably well from observations of the 21-cm line. The situation with the other major gas component, H_2 , is less clear, however, because the observations are indirect, obtained by detection of ^{12}CO and ^{13}CO . It is thought that the CO is excited by impacts with H_2 molecules and various arguments can be used to relate the CO densities to those of H_2 (see, for example, the comprehensive analysis of Sanders *et al.*¹; to be referred to as SSS).

A detailed study of the H_2 problem has been made by SSS under the usual assumption of radial symmetry (Fig. 1). This can be regarded as the 'conventional' estimate of the surface density of H_2 , $\sigma_{H_2}(R)$; here, we estimate this quantity by quite different means (see Fig. 1 legend).

The γ -ray view

In the absence of a significant contribution of γ rays from discrete sources²⁻⁴, the relation between the cosmic-ray intensity (I_{CR}), the surface density of gas and the observed γ -ray intensity is straightforward.

First, we examine very local molecular clouds where I_{CR} is presumably the same as that at the Earth and determine the best value of α , the conversion factor relating the measured quantity from CO measurements, $\int T dV$, (where T is antenna temperature and V velocity) to the column density of gas, $N(H_2)$. $\alpha = N(H_2)/\int T dV$ in units of 10^{20} atoms $cm^{-2} K^{-1} km^{-1} s$ (note that some authors use ' X ' = $\alpha/2$ in units of molecules $cm^{-2} K^{-1} km^{-1} s$). The value can then be used in the inner Galaxy, suitably modified according to our view of interstellar chemistry. Second, we use the dependence of I_{CR} on radial distance (R) in the outer Galaxy to imply its probable dependence on R in the inner Galaxy, and thereby determine the total gas surface density independently.

We start by studying the local clouds, particularly those in Orion. As well as determining a value for α , we must show conclusively that the cosmic rays are able to penetrate most of the gas in the clouds.

Local molecular clouds

Several γ -ray studies of the Orion clouds have been made already, but here we follow the analysis of Houston and Wolfendale⁵ and concentrate on the two clouds A and B. The question of a possible exclusion of cosmic rays from the dense parts of

* Permanent address: Bhabha Atomic Research Centre, Nuclear Research Laboratory, Srinagar-190006, Kashmir, India.

† Permanent address: Physics Department, Qata University, Doha, Qatar.

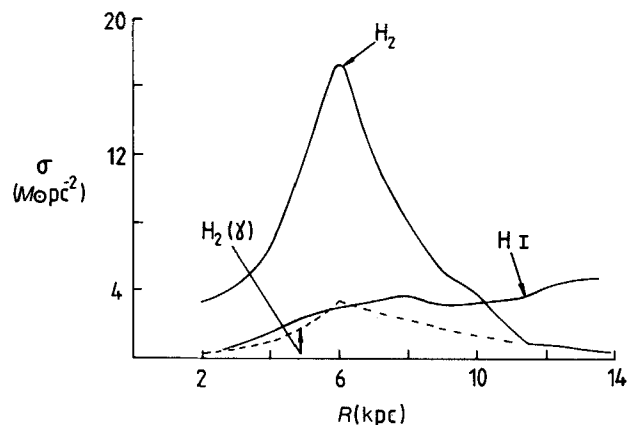


Fig. 1 Dependence of the surface density of H_2 and $H\text{ I}$ on Galactocentric radius, R . H_2 is estimated in ref. 1; in other analyses different conversion factors have been applied, resulting in maximum densities varying from about $13 M_\odot \text{ pc}^{-2}$ (ref. 8) to $32 M_\odot \text{ pc}^{-2}$ (ref. 35). The conversion factor for $^{12}\text{CO} \rightarrow H_2$ used in ref. 1 in deriving H_2 is $\alpha = 7.2 \times 10^{20} \text{ atoms cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$. The curve for $H\text{ I}$ is the mean of derivations in refs 8 and 9; both use the basic Berkeley data³⁶. $H_2(\gamma)$ is our own estimate of $\sigma_{H_2}(R)$.

the clouds has been examined⁵. Although the cosmic-ray energies are probably sufficiently high for exclusion not to be dramatic, as suggested theoretically by Skilling and Strong⁶, nevertheless an experimental demonstration is essential.

The result can be seen in Fig. 2 and its legend. Note that there is near linearity of ΔI_γ with $\int T \text{ d}V$; if linearity is assumed, α follows as the gradient of the line and is found to be $3.7 \pm 1.0 \text{ atoms cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$. This value of α is much less than that in Fig. 1 (3.7 compared with 7.2); it is in fact a little smaller than the value we used previously (for example, 4.6 in ref. 7, following Gordon and Burton⁸), although Li *et al.*⁹ made the case for $\alpha = 2.3$, locally, from various other arguments. A more detailed discussion of these α values is given later.

Figure 2 shows that there is no evidence for cosmic-ray (CR) exclusion from the densest parts of the clouds (on the angular scale of the COS B observations). If anything, there might be an excess of CR in the densest parts. If the γ -ray excess were attributed to an excess of cosmic-ray particles (or unresolved γ -ray sources), then the true value of α for the $\text{CO} \rightarrow H_2$ conversion would be significantly smaller than in Fig. 2, but the evidence is not yet strong enough for this assumption.

Other local clouds

The 'local' CR intensity adopted for Orion can probably be used out to $\sim 1 \text{ kpc}$ (despite evidence for local gradients¹⁰); the analysis of Issa and Wolfendale^{11,12} examined the local clouds. The objects in this range comprise Taurus, Corona Austrina, Ophiuchi and Gum and Monoceros OBI as well as Orion. Masses have been estimated both with and without the local thermodynamic equilibrium (LTE) condition and using $\alpha = 4.6$, and suitable allowance has been made for $H\text{ I}$ and ionized gas where appropriate. The lower limits to the mass (that is, adopting LTE) now seem more appropriate and the results of Issa and Wolfendale^{11,12} show that the average value of the CR enhancement factor (F is the ratio of I_{CR} in the cloud to I_{CR} locally) needed to satisfy the γ -ray data is $F \approx 0.5$. Thus, a re-analysis of this work gives $\alpha = 2.3$, somewhat smaller than that derived in the previous section.

Here, we have presented evidence for two local values of α , 3.7 and 2.3. Of these, the first is statistically firmer, but there is almost certainly a correction to be applied because the 'metallicity' M (defined here as the ratio of O to H, normalized to the local value) is lower in Orion than in its surroundings by a factor of ~ 1.7 (refs 13, 14).

Later, we argue that the conversion from CO to H_2 depends on the metallicity (in particular, we take α as proportional to

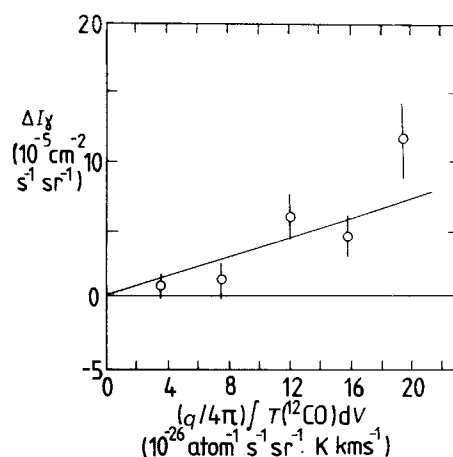


Fig. 2 Correlation of γ -ray intensity ($E_\gamma > 100 \text{ MeV}$) with $\int T(^{12}\text{CO}) \text{ d}V$ for the Orion clouds A and B (after ref. 5). ΔI_γ is the γ -ray intensity less the predicted contribution from CR interacting with $H\text{ I}$ along the line of sight. The CO data are from Kutner *et al.*³⁷ and the areas adopted were within the 1 km s^{-1} contours. The γ -ray intensities are from the COS B experiment⁴; $q/4\pi$, the emissivity per H atom, has been taken as $2.0 \times 10^{-26} \text{ atom}^{-1} \text{ s}^{-1} \text{ sr}^{-1}$, this value being the mean of various 'high latitude' estimates. The mean slope gives $\alpha = 3.7 \times 10^{20} \text{ atoms cm}^{-2} \text{ K}^{-1} \text{ km}^{-1} \text{ s}$; the errors shown are statistical.

M^{-1}); to be consistent we must adopt the same procedure with Orion to derive from its α -value the corresponding average local value α_L . Because metallicities have not been determined for other galactic molecular clouds (as distinct from $H\text{ II}$ regions in the inner Galaxy which are used to give the so-called metallicity gradient), we adopt a local average value of α higher than α/M but lower than α . Accordingly, $\alpha_L = 2.7$ is used here as a compromise.

Other estimates of α locally

The Virial Theorem can be used to give estimates of cloud masses (strictly for clouds in equilibrium). Application to the local clouds surveyed by Stark and Blitz¹⁵, after adopting intra-cloud velocities for the optically thin ^{13}CO line ($\approx 1.5 \text{ km s}^{-1}$ half-width half maximum), yields $\alpha_L \approx 2$. Direct determination of the column density of ^{12}CO from an optically thin transition in Orion¹⁶ gives an upper limit to $N(H_2)$, implying $\alpha < 3.5$, if $^{12}\text{CO}/^{13}\text{CO}$ is ≈ 40 , as often assumed. Another important result of this observation is that the velocity dispersion is $\leq 1.5 \text{ km s}^{-1}$, a small value leading again to a small virial mass and a small α . Taken together, the evidence indicates that $2 < \alpha_L < 4$, hence we adopt $\alpha_L = 2.7$.

Cosmic-ray gradient

We have demonstrated that the new value of α (2.7) is 0.37 times that used in deriving the surface density of H_2 (Fig. 1). We next examine cosmic-ray evidence for an even smaller α value in the inner Galaxy. Starting with the γ -ray method, we examine the evidence for cosmic-ray gradients, that is, systematic changes of I_{CR} with R . There is agreement amongst those who have examined the SAS II (ref. 17) and COS B (ref. 18) γ -ray results^{10,19} that there is a fall-off of CR electron intensity in the outer Galaxy (Fig. 3).

There is an approximate agreement of I_{CR} with the surface density of supernova remnants (Fig. 3) at the lower energies. There are admittedly problems with the determination of I_{CR} in the inner Galaxy (because of uncertainty in the effect of allowing for the ubiquitous H_2 component), but the situation in the outer Galaxy is well established. Here, there is fall-off in I_{CR} (ref. 10) and it is probable on general grounds that there will be a corresponding increase in I_{CR} in the inner Galaxy. It would be surprising if this increase in I_{CR} did not occur to at

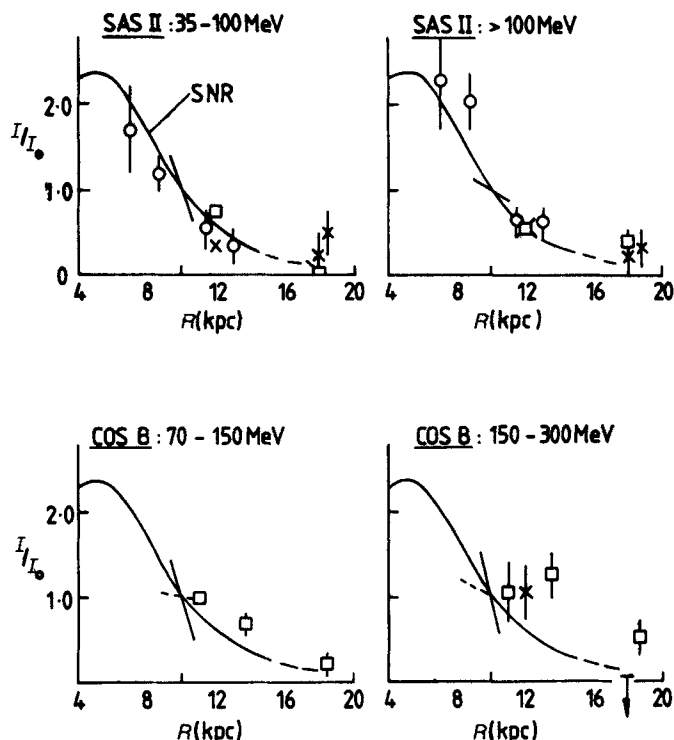


Fig. 3 Gradient of cosmic-ray intensity derived by various workers from the data of the SAS II experiment¹⁷ and COS B¹⁸. The ordinate is the ratio of the cosmic-ray intensity to that locally, I/I_0 . The SNR distribution³⁸ is shown as a probable distribution of the low energy cosmic rays responsible for the γ -rays. $\circ$, Data from Issa *et al.*³⁹; $\square$, data from Bloemen *et al.*¹⁹; $\times$, data from Bhat *et al.*^{10,40}. The inclined lines through $R = 10$ kpc (the local position) refer to an analysis of γ -rays in the range $|b| = 11-19^\circ$ (ref. 40); dotted inclined lines refer to an extreme Inverse Compton contribution; errors are statistical.

least $R = 6$ kpc (all other indicators of galactic activity, such as H II regions, pulsars and H_2 surface density show similar increases). The effect of such an increase on the required surface density of all galactic gas is considered below.

Surface density of gas

The estimation of the surface density of gas at $R = 6$ kpc as required by the γ -ray data is straightforward. Figure 4 gives the result. Taking the SSS value of $\sigma_{H_2}(R = 6 \text{ kpc})$ as datum and 'correcting' to our new value of α (that is, $2.7/7.2 = 0.37$ times) yields a value of σ_{H_2} still too high for conformity with the γ -ray results; a further reduction is required. Therefore, we investigate the metallicity in the ISM and introduce a 'metallicity correction'^{20,21}. Briefly, the O/H ratio in the ISM increases towards the Galactic Centre (GC) and we consider this to cause an increase in the $^{12}\text{CO}/H_2$ conversion, the factor of increase amounting to two at $R = 6$ kpc and $\sim 3-5$ at the GC. Applying the factor two, there is now a reduction in $\sigma_{H_2}(R = 6 \text{ kpc})$ by a factor $0.37 \times 0.5 = 0.18$ of the SSS value, in agreement with the γ -ray expectation.

The use of the metallicity correction is controversial; it is argued against in ref. 1. The arguments for the correction are described in ref. 21 and further evidence is obtained from more detailed studies of interstellar chemistry described below. Higher temperatures and probable greater shock density in the inner Galaxy may augment or simulate a metallicity dependence.

Interstellar chemistry

We investigate the relationship between $N(^{12}\text{CO})$ (the apparent column density of ^{12}CO), the actual value of N_{H_2} and the metallicity, characterized by the O/H ratio. We have assumed here that α is proportional to M^{-1} and we seek support for this

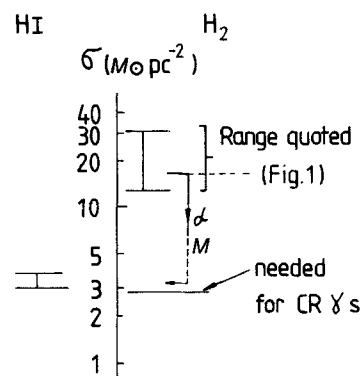


Fig. 4 Surface densities of H I and H_2 at $R = 6$ kpc. The range of quoted values from CO observations (using various α values) is indicated, as is the value from Fig. 1. The effect of reducing the surface density of H_2 by 0.37 times (to allow for the change in α) and by 0.5 (to allow for metallicity) is shown. Also indicated is the value needed to satisfy the cosmic γ -ray results.

relationship from arguments of interstellar chemistry. Comprehensive calculations have been made recently (D. A. Williams and T. Millar, personal communication and ref. 22).

Starting with the important optically-thin ^{13}CO component and limiting the argument to spherical inert clouds of mass $M_C (\sim 10^4 M_\odot)$, the column density along a diameter divided by N_{H_2} varies approximately as $M^{1.2}$ (specifically for $T = 10$ K, gas density $n = 600$ nuclei cm^{-3}), that is, α is proportional to $M^{-1.2}$. The result is very similar for all densities in the range $200 < n < 1,800 \text{ cm}^{-3}$ and does not depend on cloud mass for $3 \times 10^3 < M_C < 10^5 M_\odot$. The dependence on the ultraviolet intensity and the cosmic-ray heating rate is slight.

The results with ^{12}CO are very similar, although because ^{12}CO is optically thick, the detected signal in practice would not depend necessarily on M in the same way. It is here, however, that we return to the fact that over most of the Galaxy there is proportionality between the strengths of the ^{12}CO and ^{13}CO signals; we interpret this in terms of turbulence in the clouds and associated density variations such that velocity smearing allows ^{12}CO to be detected over wide regions of clouds. The result is that, although the premise of large uniform clouds (on which the chemistry calculations were made) breaks down, it seems to be in the sense of increasing the metallicity rather than reducing it. More sophisticated cloud modelling is necessary, but the best estimate so far is that α is approximately proportional to M^{-1} .

Inner Galaxy α estimates

There are other ways of estimating α at $R \approx 6$ kpc, notably involving the virial mass, X-ray and infrared absorption, which are summarized here and discussed elsewhere in detail.

The virial masses are low when velocity dispersions for optically thin lines are used and we again have $\alpha \approx 2$. The X-ray technique, which measures essentially the column density of atoms such as O, was first used with success by Blitz and Shu²⁰. We have updated the treatment using new observations, with the result that $\alpha \approx 1$ for $R \approx 6$ kpc. Infrared observations give a complementary view; using the extinction data of Eaton *et al.*²³ we again find $\alpha \approx 1$. The dust/gas ratio is probably lower by a factor of two in the region $R \approx 6$ kpc (ref. 24) and applying this factor increases α to two.

The range of the α -values is 2 (virial), 1 or 2 (infrared) or 1 (X-ray), consistent with the γ -ray result, where $\alpha \approx 1.4$.

Galactic Centre

Indirect evidence about the $\text{CO} \rightarrow H_2$ calibration comes from the claim by millimetre wave astronomers for considerable masses of H_2 within ~ 0.4 kpc of the GC on the basis of high

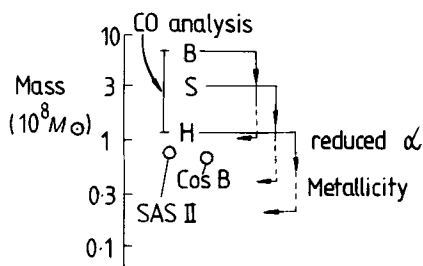


Fig. 5 Mass of gas within 400 pc of the Galactic Centre. The range of values corresponds to different CO analyses: B, data from Bania⁴¹; S, from Sanders *et al.*¹; H, data from Heiligman⁴². Using our own α and the metallicity correction ($M=3$) gives the reduced values shown by the arrows. Circles represent masses derived from the SAS II and COS B data assuming that the CR intensity at the Galactic Centre is the same as that locally (that is, $F=1$). Adopting the mean of the reduced mass estimates would indicate $F \approx 2$, a reasonable result.

$\int T(^{12}\text{CO}) dV$ values in that direction. Several estimates have been made (Fig. 5).

For the γ -ray fluxes from the GC region, we follow the recent analysis of Houston and Wolfendale²⁵, who derived a GC 'source' flux of $2.3 \times 10^{-6} \text{ cm}^{-2} \text{ s}^{-1}$ for $E > 100 \text{ MeV}$ for SAS II. For COS B, we identify the source 2CG359-00 from the catalogue of Hermesen²⁶ with the GC; this has a flux above the same energy of $1.8 \times 10^{-6} \text{ cm}^{-2} \text{ s}^{-1}$.

The problem now is the expected magnitude of I_{CR} in the GC region. A detailed study by Wolfendale and Worrall²⁷ shows that a self-consistent model can be made in which the production rate of CR per unit mass of gas is approximately the same at the GC and locally (and is thus $\approx 100 \times$ greater than locally per unit volume at the GC). In the model, there is agreement with radio synchrotron data in showing that the ambient I_{CR} at the GC is $\sim 10 \times$ that near the Sun. If this factor were used, we would expect a GC mass of $\approx 0.07 \times 10^8 M_{\odot}$, a very low value. As a lower limit to I_{CR} at the GC the local value can be used (that is $F=1$), resulting in a mass of $\approx 0.7 \times 10^8 M_{\odot}$. The masses shown in Fig. 5 use $F=1$.

The effect of applying the new α and the metallicity correction (a reduction of at least $M=3$ at the GC) can now be examined. Figure 5 shows the new CO-implied masses which are now 'reasonable' in that they correspond to a modest CR enhancement of intensity ($F \approx 2$). With the higher metallicity factor $M=5$, implied by extrapolating the O/H gradient of -0.07 kpc^{-1} back to the GC, the mass would be correspondingly lower and $F \approx 4$, a still reasonable enhancement.

Blitz *et al.*²⁸ have recently made a similar analysis and concluded that a very small α at the GC is probable. The main value of the present analysis of the GC is that it lends support to the presence of a metallicity effect (or perhaps shocks and temperature effects simulating it) and thereby to the correctness of our α -values in the inner Galaxy. (Blitz *et al.*²⁸ also make important comments about 'time variability' of the CO results; that is, the presence of systematic errors between different instruments and sets of measurements.)

Discussion

We have derived an independent estimate of the magnitude of the surface density of gas in the Galaxy as a function of galactocentric radius using γ -ray astronomy and supported by other techniques. Figure 1 shows that our estimate, $H_2(\gamma)$, is considerably lower than the 'conventional' estimate, from ref. 1, an estimate made using the usual procedure of relating $\int T dV$ for the 2.6-mm CO line to the column density of H_2 . Although we argue that the new estimate is reliable, it also contains assumptions; therefore, we have examined the 'conventional' method and, using the basic data on CO, made our own estimate of the surface density of the molecular gas.

The basic results¹ for $\int T dV$ versus longitude have been taken, together with our own unfolding, to give the radial distribution

of CO. Adopting radial symmetry, the result is CO surface densities ~ 60 – 80% of those given by SSS. This discrepancy presumably relates to differing assumptions in allowing for the clumpy nature of the gas distribution.

The important question of the $\text{CO} \rightarrow \text{H}_2$ conversion has been examined in detail. Dickman's²⁹ results of $\int T(^{13}\text{CO}) dV$ against visual extinction A_V for local clouds, and similar data for Taurus³⁰, have been examined and, as found by SSS, there is evidence for a saturation of $\int T dV$ versus A_V . ($\int T dV$ is proportional to $A_V^{0.5}$, approximately fitting a power law to the whole range in ref. 1, but we can have $\int T dV$ proportional to A_V up to values of 6 for A_V for Taurus at least, with a flattening above.) It is conventional to follow Bohlin *et al.*³¹ and adopt a linear relation of N_{H_2} versus A_V ($N_{\text{H}_2} = 0.95 \times 10^{21} A_V \text{ molecules cm}^{-2}$, although this relation has only been verified observationally for $A_V < 2$, clearly another region of uncertainty). The crucial questions concern the effective mean value of $\int T dV$ for the GMC, where on the $A_V^{0.5}$ plot are we situated and the corresponding α value. These questions are examined in detail below.

For the Orion molecular clouds the 'median mass' $\int T(^{12}\text{CO}) dV$ value (half the cloud mass is within and half without) is $\sim 11 \text{ K km s}^{-1}$ and 2 K km s^{-1} for ^{13}CO . Such a value is well on the linear part of the $\int T(^{13}\text{CO}) dV$ against A_V relation.

Inspection of the data on Dickman's²⁹ clouds, five of which are completely mapped, indicates that the mean value of A_V is ~ 2.5 and the median $\int T(^{13}\text{CO}) dV \sim 4 \text{ K km s}^{-1}$. It is immediately apparent that with such a low value of $\int T(^{13}\text{CO}) dV$, the derived α value will be low; specifically 3.0 for the five fully mapped clouds, very close to our earlier γ -ray value of 2.7 for the local region.

There is evidence that the 'clouds' detected on the galactic plane in the inner Galaxy have high values of $\int T(^{13}\text{CO}) dV$ (for example, ref. 32, 7.2 K km s^{-1} , used in ref. 1), although other data³³ give values of only $\sim 3 \text{ K km s}^{-1}$. We consider that many of the so-called clouds are almost certainly conglomerates of smaller clouds, each of the latter probably resembling those seen locally (Orion, Taurus). If this is so, we are not on the saturated part of $\int T dV$ versus A_V , which corresponds for local clouds to those rare passages through very dense regions, but, usually to low values of $\int T dV$, on the linear part. For example, with passage through three clouds, the operative value of $\int T dV$ is $7.2/3 = 2.4 \text{ K km s}^{-1}$. This contention is supported by data on identified clouds of large radius³⁴, indicating that the mean density (molecules cm^{-3}) of the clouds falls as their linear dimension increases; this is what we would expect if loose conglomeration were occurring. The point now is that the separate parts of the conglomerate have different velocities and the constituent $\int T dV$ values are additive.

Using the analysis presented here, we can derive the 'new' mass of H_2 in the inner Galaxy (corresponding to $H_2(\gamma)$ in Fig. 1). Neglecting the region $R < 1 \text{ kpc}$, the value is $6 \times 10^8 M_{\odot}$, compared with $8 \times 10^8 M_{\odot}$ for H I.

We thank Drs M. G. Edmunds, L. Blitz and T. Millar and Professors B. E. J. Pagel, M. Peimbert and D. A. Williams for helpful comments and discussions.

Received 2 July 1984; accepted 8 February 1985.

- Sanders, D. B., Solomon, P. M. & Scoville, N. Z. *Astrophys. J.* **276**, 182–203 (1984).
- Fichtel, C. E. & Kniffen, D. A. *Astr. Astrophys.* **134**, 13–23 (1983).
- Houston, B. P. & Wolfendale, A. W. *J. Phys.* **G10**, 1587–1598 (1984).
- Houston, B. P. thesis, Univ. Durham (1984).
- Houston, B. P. & Wolfendale, A. W. *J. Phys. G* (in the press).
- Skilling, J. & Strong, A. W. *Astr. Astrophys.* **53**, 253–258 (1976).
- Arnaud, K. *et al. Mon. Not. R. astr. Soc.* **201**, 745–758 (1982).
- Gordon, M. A. & Burton, W. B. *Astrophys. J.* **208**, 346–353 (1976).
- Li, T. pei, Riley, P. A. & Wolfendale, A. W. *Mon. Not. R. astr. Soc.* **203**, 87–102 (1982).
- Bhat, C. L., Mayer, C. J. & Wolfendale, A. W. *Astr. Astrophys.* **140**, 284–287 (1984).
- Issa, M. R. & Wolfendale, A. W. *Nature* **292**, 430–433 (1981).
- Issa, M. R. & Wolfendale, A. W. *J. Phys. G*, **7**, L187–L191 (1981).
- Pagel, B. E. J. & Edmunds, M. G. *A. Rev. Astr. Astrophys.* **19**, 77–113 (1981).
- Peimbert, M. *Ann. N.Y. Acad. Sci.* **395**, 24–31 (1982).
- Stark, A. A. & Blitz, L. *Astrophys. J.* **225**, L15–L19 (1978).
- Black, J. H. & Willner, S. P. *Astrophys. J.* **279**, 673–678 (1984).
- Fichtel, C. E. *et al. NASA Tech. Mem.* X-662-76-198 (GSFC Greenbelt, 1976).
- Mayer-Hasselwander, H. A. *et al. Astr. Astrophys.* **105**, 164–175 (1982).
- Blomen, J. E. G. M. *et al. Astr. Astrophys.* **135**, 12–22 (1984).
- Blitz, L. & Shu, F. H. *Astrophys. J.* **238**, 148–157 (1980).
- Li, T. P., Riley, P. A. & Wolfendale, A. W. *J. Phys. nucl. Phys.* **G8**, 1141–1154 (1982).

22. Williams, D. A. & Hartquist, T. W. *Mon. Not. R. astr. Soc.* **210**, 141-145 (1984).
23. Eaton, N., Adams, D. J. & Giles, A. P. *Mon. Not. R. astr. Soc.* **208**, 241-251 (1984).
24. Bhat, C. L., Mayer, C. J. & Wolfendale, A. W. (in preparation).
25. Houston, B. P. & Wolfendale, A. W. *Astr. Astrophys.* **126**, 22-30 (1983).
26. Hermesen, P. *Phil. Trans. R. Soc. A301*, 519-521 (1981).
27. Wolfendale, A. W. & Worrall, D. M. *Astr. Astrophys.* **60**, 165-170 (1977).
28. Blitz, L., Bloemen, J. B. G. M., Hermesen, W. & Bania, T. M., *Astr. Astrophys.* (in the press).
29. Dickman, R. L. *Astrophys. J. Suppl.* **37**, 407-427 (1978).
30. Freking, M. A., Langer, W. D. & Wilson, R. W. *Astrophys. J.* **262**, 590-605 (1982).
31. Bohlin, R. C., Savage, B. D. & Drake, J. F. *Astrophys. J.* **224**, 132-142 (1978).
32. Liszt, H. S., Xiang, D. & Burton, W. B. *Astrophys. J.* **249**, 532-549 (1981).
33. Stark, A. A. thesis, Univ. Princeton (1979).
34. Dame, T. M. thesis, Goddard Inst. for Space Studies, New York (1984).
35. Solomon, P. M. & Sanders, D. B. in *Giant Molecular Clouds in the Galaxy* (eds Solomon, P. M. & Edmunds, M. G.) 41-73 (Pergamon, Oxford, 1980).
36. Weaver, H. & Williams, D. R. W. *Astr. Astrophys. Suppl.* **8**, 1-503 (1973).
37. Kutner, M. L., Tucker, K. D., Chin, G. & Thaddeus, P. *Astrophys. J.* **215**, 521-528 (1977).
38. Kodaira, K. *Publ. astr. Soc. Jap.* **26**, 255-265 (1974).
39. Issa, M. R., Riley, P. A., Strong, A. W. & Wolfendale, A. W. *J. Phys. G*, **7**, 973-994 (1981).
40. Bhat, C. L., Houston, B. P., Issa, M. R., Mayer, C. J. & Wolfendale, A. W. *Proc. Abingdon ISM Conf.* 39-54 (1984).
41. Bania, T. M. *Astrophys. J.* **216**, 381-403 (1977).
42. Heiligman, G. M. thesis, Univ. Princeton (1982).

LETTERS TO NATURE

Acceleration of cosmic rays in the Loop I 'supernova remnant'?

C. L. Bhat*, M. R. Issa†, C. J. Mayer
& A. W. Wolfendale

Physics Department, University of Durham, Durham DH1 3LE, UK

Attempts are being made to establish the origin of cosmic rays using the γ -ray method, that is, by dividing γ -ray intensity by gas-column density to chart the cosmic-ray intensity. Here we present evidence from satellite data on γ rays and from new results on the distribution of gas in the interstellar medium (ISM) for an enhancement of electron and proton intensity in the Loop I supernova remnant (SNR). There follows from this a straightforward explanation for the origin of the bulk of the low-energy cosmic radiation.

Here, we describe a study of the secondary γ rays in the 100-MeV region, generated by the interactions of cosmic rays (protons of 1-10 GeV and electrons of several hundred MeV) with the ISM. It is generally considered^{1,2} that most of the γ rays detected in the satellite experiments are, in fact, generated by these interactions rather than by unresolved discrete sources, particularly at the galactic latitudes, $|b| > 10^\circ$, considered here.

Early work by Dodds *et al.*³ using SAS II data⁴ gave evidence (subsequently confirmed^{5,6}) for a gradient of cosmic-ray intensity in the outer galaxy, indicative of a galactic origin for the mixed beam of relevant electrons and protons. To distinguish between the relative contributions of protons and electrons, however, is not a trivial problem and, although it is now generally agreed that there is a gradient for electrons^{5,6}, the situation for protons is debatable^{6,7}. (In fact, acceptance of the arguments presented here would resolve the debate; we shall argue that the strong local gradient⁷ will have an important contribution from the Loop.)

That there is a strong correlation of the γ -ray flux with the intensity of synchrotron radiation, both in and out of the plane is well known^{8,9}. Prominent features of the synchrotron map out of the plane are a number of Loops of which the most prominent is Loop I. This has been identified by Berkhuysen *et al.*¹⁰ with a SNR centred on a point ~ 130 pc away, with diameter ~ 230 pc and having central coordinates $l = 329^\circ$, $b = +17.5^\circ$. The corresponding position for the supernova was 40 pc above the plane and the bulk of the shock front is still in the gas disk; the Loop is prominent between latitudes -30° to $+80^\circ$ and longitudes 260° to 60° .

It is natural to search for a specific correlation of the γ -ray flux with this Loop to support a popular hypothesis¹¹⁻¹⁵ that the supernova is the seat of cosmic-ray acceleration. We mention particularly the work of Blandford and Cowie¹⁵, who have not only considered the cosmic-ray energetics of radio loops but have made specific predictions of γ -ray emissivities.

Concerning γ rays, there are data from the SAS II experiment¹⁶ and, for each of a number of latitude bands (Fig. 1), we

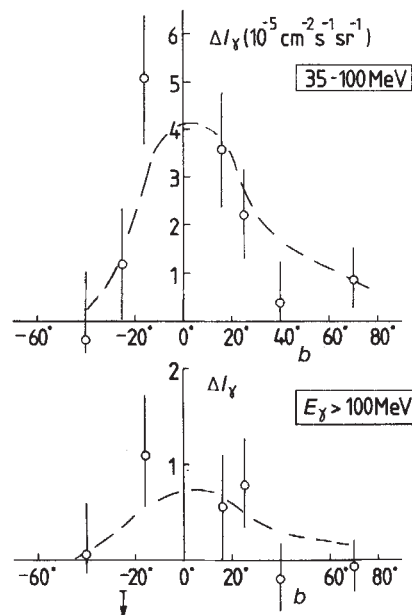


Fig. 1 The excess γ -ray intensity associated with Loop I as a function of galactic latitude. The basic γ -ray data are from the SAS II experiment¹⁶. ΔI_γ is the difference between observed and expected γ -ray intensity for the Loop region minus the same quantity outside the Loop. The smooth curve is the corresponding T_{408} , averaged in the same way, normalized in height to observation. The particles generating the γ rays are almost entirely electrons in the lower energy band. At higher energies protons generate about 60% of the γ rays.

have determined the dependence of γ -ray intensity on longitude. The expected γ -ray intensity for the same conditions has been derived assuming a constant cosmic-ray intensity and adopting a new estimate of the column density of gas.

The column density of gas (N) comprises two components— $H I$, N_{HI} and H_2 , N_{H_2} . The former is well known¹⁷⁻¹⁹, but the latter is uncertain and it is on the choice for N_{H_2} that the present work rests; thus, if we have seriously underestimated N_{H_2} , our conclusions will be invalid. Our method consists of using galaxy counts²⁰ to give N_{H_2} for those regions where CO observations are not available and calibrating the galaxy-count column density against N_{H_2} in the region where CO data are available ($l: 5^\circ$ – 35° ; $b: 10^\circ$ – 20° ; ref. 21). For the longitude range above $l = 210^\circ$, where galaxy counts cease, we adopt interstellar reddening data^{8,9}. As an indication of the importance of H_2 in the lowest latitude range considered (12.8° – 19.5°), the adopted average N_{H_2} is $\sim 25\%$ that for N_{HI} . Other spot measurements using different techniques in the l, b range considered also give very small N_{H_2} values^{22,23}. For latitudes outside $|b| = 20^\circ$, we follow previous work^{24,25}, neglecting N_{H_2} altogether. The emissivity values used to derive the predicted γ -ray intensities are $q/4\pi = 3.7 \times 10^{-26}$ and 2.6×10^{-26} atom $^{-1}$ s $^{-1}$ sr $^{-1}$ for 35–100

Permanent addresses: * Bhabha Atomic Research Centre, Srinagar, India; † University of Doha, Qatar.

and >100 MeV respectively. (In fact the absolute values are not very important because of the subtraction technique used.)

The procedure is now to determine the excess of the observed γ -ray intensity relative to that expected as a function of longitude, $I_e(l)$, both inside the loop region, $I_{e,L}$, and outside it, $I_{e,o}$, and to determine $\Delta I_\gamma = I_{e,L} - I_{e,o}$, representing the excess contributed by the Loop (the longitude range embraced by the Loop is $l = 260^\circ \rightarrow 60^\circ$ over the relevant latitudes). Figure 1 shows the Loop contribution as a function of latitude. A similar analysis has been made for the COS B data²⁰ in the latitude range $|b| = 11^\circ \rightarrow 19^\circ$, but, unfortunately, no COS B data are yet available for $|b| > 19^\circ$. Figure 2 gives a comparison of the Loop contributions for the positive (N) and negative (S) latitude ranges ($11^\circ \rightarrow 19^\circ$) (correcting the SAS II values to the same b -range).

There is no question of the γ -ray excess being constant over the SNR; in fact, it peaks just inside the well known radio-continuum North Polar Spur, but, insofar as we wish to determine the energetics of the remnant, the analysis is made in terms of averages over the object. Figure 1 contains a plot of ΔT_{408} (T_{408} is the antenna temperature of the radio emission at 408 MHz) against latitude from which the general correlation can be seen. It is obvious that there is clear evidence for a Loop contribution, particularly for the 35–100 MeV data.

Discounting the possibility that a dense gas cloud in the direction of the Loop has been missed, the most probable alternative explanation to cosmic-ray acceleration in a SNR is that we are dealing with inverse Compton (IC) interactions of electrons with an enhanced radiation field in the vicinity of the Loop. Using the analysis of Riley and Wolfendale⁸, it seems unlikely that $>50\%$ of the Loop contribution can be attributed to IC, at any energy; Bloemen's work would give an even smaller IC contribution (J. B. G. M. Bloemen, personal communication). It is reassuring that the COS B excess is present at all energies and, given that cosmic-ray protons predominate as primaries of the γ rays in the highest energy region, this is the best evidence for an excess of protons in the Loop.

We turn now to questions of enhanced cosmic-ray intensities and energetics, starting with the electron component. That the Loop is actually seen in synchrotron radiation shows that there is an enhanced cosmic-ray electron intensity and/or an increased magnetic field intensity (the electron energy being ~ 3 GeV for typical ISM fields). Surprisingly perhaps the magnetic field seems not to be enhanced²⁶ and thus the electron intensity must be increased considerably. Heiles *et al.*²⁶ estimate an increase by a factor $F_e \sim 20$ in the radio-continuum North Polar Spur. Averaging over the whole SNR gives an electron enhancement factor $F_{e,r} \sim 5$ –10.

Figure 1 ($E_\gamma = 35$ –100 MeV) can be used to estimate F_e for the initiating electrons for the γ rays ($E_e \sim 200$ MeV). Using the model of Blandford and Cowie¹⁵ with a mean density of 0.4 cm^{-3} (0.2 cm^{-3} in the hot ISM and 0.2 cm^{-3} for the average contribution of cold sub-clouds) the average enhancement factor is $F_{e,\gamma} \sim 6$, in good agreement with $F_{e,r}$.

We now consider the ambient energy density of the cosmic-ray electron component needed to calculate the electron energy residing in the remnant. The magnitude seems to depend critically on the lower limit to the energy, unlike the proton case, because of the need for a steep electron spectrum below 1 GeV to explain the ambient γ -ray and radio spectra. Adopting the interstellar electron spectrum of Strong and Wolfendale²⁷, the energy densities in electrons are $2.3 \times 10^{-2} \text{ eV cm}^{-3}$ above 100 MeV and $3 \times 10^{-3} \text{ eV cm}^{-3}$ above 1 GeV, to be compared with $\sim 0.5 \text{ eV cm}^{-3}$ above 100 MeV and $\sim 0.4 \text{ eV cm}^{-3}$ above 1 GeV for protons.

Adopting the 100-MeV limit and $F_{e,\gamma} = 6$, the total energy in electrons in the remnant is $\sim 2 \times 10^{49} \text{ erg}$. Turning now to protons, the mean value for SAS II and COS B is $F_{p,\gamma} \sim 4$ and the corresponding energy is $2 \times 10^{50} \text{ erg}$. Considering the problem of the IC contribution, it is best to reduce these values by a factor of two, to $E_e \sim 10^{49} \text{ erg}$ and $E_p \sim 10^{50} \text{ erg}$ and attach an uncertainty of a factor of three to each.

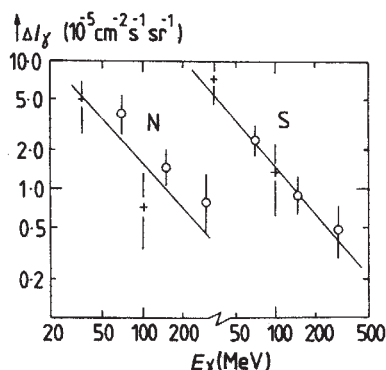


Fig. 2 The excess γ -ray intensity defined in Fig. 1 as a function of threshold energy. The latitude ranges are: N $11^\circ \rightarrow 19^\circ$, S $-11^\circ \rightarrow -19^\circ$. The circles refer to COS B data and the crosses to SAS II. The spectral shape is not dissimilar to the average spectrum of galactic γ rays. There appears to be a significant excess for the highest energy γ rays, these quanta being generated very largely by cosmic-ray protons.

We can now make a comparison with the predictions of Blandford and Cowie¹⁵; their average γ -ray surface brightness, above 100 MeV, of $7 \times 10^{-6} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ is very similar to that in Fig. 2. Concerning the energy content of the Loop, our estimate is $\sim 10^{50} \text{ erg}$ for protons and $\sim 10^{49} \text{ erg}$ for electrons, again close to estimates based on the Blandford and Cowie model for a supernova of initial energy 10^{51} erg .

We can now return to the important question of the acceleration of protons; unlike electrons, for which there is the synchrotron enhancement, the γ -ray results alone provide supporting evidence. The argument in favour of accelerated protons is demonstrated by the higher energy band in Fig. 1 (protons probably give $\sim 60\%$ of the γ rays) and the two highest energy bands in Fig. 2, particularly that for $E_\gamma > 300$ MeV, where we expect at least 80% to have been produced by protons. An important feature here is that the correlation between I_γ (> 300 MeV) and T_{408} is singularly strong, thus confirming that the protons are being accelerated along with the electrons. That we find approximate agreement between $F_{e,r}$ and $F_{e,\gamma}$ suggests that our estimates of gas column density are not seriously in error.

The implications of the results for the origin of cosmic rays are obvious, the most important feature being that the efficiency for conversion of supernova energy into cosmic-ray protons is very high. For Loop I we have presently about 10% of the initial energy in protons, which presumably will soon be released into the ISM. The energy requirement for cosmic radiation in the Galaxy can be calculated as follows. With a galactic volume for cosmic rays of $8 \times 10^{67} \text{ cm}^3$, a lifetime for the particles of $2 \times 10^7 \text{ yr}$ and an ambient energy density of 0.5 eV cm^{-3} , the cosmic-ray energy input is $\sim 10^{41} \text{ erg s}^{-1}$. Adopting a supernova rate of 1 per 30 yr and total energy per supernova of 10^{51} erg , the efficiency required for generation of cosmic radiation at the moment of release into the ISM is simply 10%, that is, the value we estimate for Loop I. Such agreement is surely fortuitous in view of the many uncertainties involved, but it does demonstrate the cogency of the entire argument. We are currently looking for the expected cosmic-ray enhancements in other SNR to attempt to confirm our conclusions.

We thank the SERC for its support and Dr J. L. Osborne, Dr L. Blitz and Mr M. Rogers for helpful discussions.

Received 7 December 1984; accepted 15 February 1985.

1. Fichtel, C. E. & Kniffen, D. A. *Astr. Astrophys.* **134**, 13–23 (1984).
2. Houston, B. P. & Wolfendale, A. W. *Astr. Astrophys.* **126**, 22–30 (1983); *J. Phys. G.* (in the press).
3. Dodds, D., Strong, A. W. & Wolfendale, A. W. *Mon. Not. R. astr. Soc.* **171**, 569–577 (1975).
4. Fichtel, C. E. *et al. Goddard Space Flight Center X-662*, 74–304 (1974); *Astrophys. J.* **198**, 163–182 (1975).
5. Issa, M. R., Riley, P. A., Strong, A. W. & Wolfendale, A. W. *J. Phys. G.* **7**, 973–994 (1981).
6. Bloemen, J. B. G. M. *et al. Astr. Astrophys.* **135**, 12–22 (1984).
7. Bhat, C. L., Mayer, C. J. & Wolfendale, A. W. *Astr. Astrophys.* **140**, 284–287 (1984).
8. Riley, P. A. & Wolfendale, A. W. *J. Phys. G.* **10**, 1149–1161 (1984).
9. Haslam, C. G. T., Kearsey, S., Osborne, J. L., Phillips, S. & Stoffel, H. *Nature* **289**, 470–473 (1983).

10. Berkhuijsen, E. M., Hasam, C. G. T. & Salter, C. J. *Astr. Astrophys.* **14**, 252-262 (1971).
11. Zwicky, F. *Proc. natn. Acad. Sci.* **25**, 338-344 (1939).
12. Ginzburg, V. L. & Syrovatsky, S. I. *The Origin of Cosmic Rays* (Pergamon, Oxford, 1964).
13. Axford, W. I. *Ann. N.Y. Acad. Sci.* **375**, 297-313 (1981).
14. Drury, L. O'C. & Volk, H. J. *Astrophys. J.* **248**, 344-351 (1981).
15. Blandford, R. D. & Cowie, L. L. *Astrophys. J.* **260**, 625-634 (1982).
16. Fichtel, C. E. *et al. NASA Tech. Memo.* 79650 (1978).
17. Weaver, H. & Williams, R. W. *Astr. Astrophys. Suppl.* **8**, 1-503 (1973).
18. Heiles, C. & Habing, H. J. *Astr. Astrophys. Suppl.* **14**, 1-50 (1974).
19. Heiles, C. & Cleary, M. N. *Aust. J. Phys. Astrophys. Suppl.* **47**, 1-58 (1979).
20. Strong, A. W. *et al. Astr. Astrophys.* **115**, 404-412 (1982).
21. Lebrun, F. & Huang, Y. L. *Astrophys. J.* **281**, 634-638 (1984).
22. Bohlin, R. C., Savage, B. D. & Drake, J. F. *Astrophys. J.* **224**, 132-142 (1978).
23. Blitz, L., Fich, M. & Stark, A. A. *Astrophys. J. Suppl.* **49**, 183-206 (1982).
24. Dewey, R. J., Uson, J. M., Weisberg, J. M., Wilkinson, D. T. & Stark, A. A. *Astr. J.* **88**, 1832-1834 (1983).
25. Blitz, L., Magnani, L. & Mundy, L. *Astrophys. J.* (in the press).
26. Heiles, C., Chu, Yuo-Hua, Reynolds, R. J., Yegingil, I. & Troland, T. H. *Astrophys. J.* **242**, 533-540 (1980).
27. Strong, A. W. & Wolfendale, A. W. *J. Phys. G.* **4**, 1793-1807 (1978).

Terrestrial impactors at geological boundary events: comets or asteroids?

Paul R. Weissman

Earth and Space Sciences Division, Jet Propulsion Laboratory,
4800 Oak Grove Drive, Pasadena, California 91109, USA

Evidence has been presented^{1,2} for a 26-28-Myr periodicity in both the terrestrial extinction record and the age of large well-dated impact craters on the Earth. A cometary source controlled by perturbations associated with either the Solar System's galactic z-motion, which has a 33-Myr half-period, or an unseen solar companion star in a 26-28-Myr orbit has been suggested³⁻⁵ for the impacts. Identifications of meteoritic material in the impact melts of the craters⁶ used to demonstrate the periodicity, however, are indicative of highly differentiated impactors, which are more representative of asteroids than of comets. Clustering of crater ages at some dates may be indicative of random comet showers, some associated with extinctions and some not. The source of asteroidal impactors with a ~28-Myr period remains unknown and casts doubt on the entire periodicity argument.

The cause of biological extinction events at geological boundaries has been much studied recently, especially since the detection of extraterrestrial material at the Cretaceous-Tertiary (K-T) boundary⁷. Among the extinction triggering mechanisms suggested are passage of the Solar System through interstellar dust clouds, asteroid and comet impacts, geomagnetic field reversals, nearby supernova explosions and exposure of the Earth to intense galactic cosmic rays. The impact arguments have gained wide acceptance, at least for the well-studied K-T event where shocked quartz grains (presumably impact ejecta) have been found⁸, in addition to the large global siderophile anomalies.

The 28-Myr periodicity in the crater record is based on 16 craters with diameters >5 km and ages whose probable error is <20 Myr^{2,9}. These craters are listed in Table 1 along with their diameters and ages. For eight of the sixteen craters, identifications of the impacting body composition have been made (Table 1) based on meteorite siderophile signatures in the impact melt⁶; in some cases the identification of impactor types is not completely certain.

Of the eight identified impactor compositions, six are differentiated or highly differentiated meteorite types. Formation of iron, stony and achondrite meteorites all require melting and differentiation in fairly large parent bodies. Ordinary chondrites may or may not be partially melted, but are depleted in volatiles as compared with the more-primitive carbonaceous chondrites. Only the impactors at Wanapitei, Canada at 37±2 Myr and Lappajarvi, Finland at 77±4 Myr are primitive undifferentiated meteorite types. Interestingly, Lappajarvi is the only crater in the list (except for the youngest one, of age 7 Myr) which is not identified with a periodic impact²; it falls almost exactly between the impact cycles at ~65 and 95 Myr.

Table 1 Periodic terrestrial craters

Location	Diameter* (km)	Age (Myr)*	Projectile type†
Karla, USSR	10	7±4	
Haughton, Canada	20	13±11	
Ries, Germany	24	14.8±0.7	Achondrite (?)
Mistastin, Labrador	28	38±4	Iron or achondrite (?)
Wanapitei, Ontario	8.5	37±2	Chondrite (C1, C2, LL)
Popigai, Siberia	100	39±9	Iron
Lappajarvi, Finland	14	77±4	C-chondrite
Steen River, Alberta	25	95±7	
Boltys, Ukraine	25	100±5	
Logoisk, USSR	17	100±20	
Mien Lake, Sweden	5	118±2	Stony (?)
Gosses Bluff, Australia	22	130±6	
Rochechouart, France	23	160±5	Iron (IIA) or chondrite (?)
Obolon, Ukraine	15	185±10	Iron
Puchezh-Katunki, USSR	80	183±3	
Manicougan, Quebec	70	210±4	

* Data from refs 2 and 6.

† Data from refs 6 and 9.

It is difficult to reconcile impact craters formed by highly differentiated meteorite types with a cometary source. Comets are believed to have formed¹⁰ as primitive undifferentiated icy planetesimals in the region of the outer planets and beyond. The typical cometary nucleus is believed to be a heterogeneous mixture of icy volatiles (mostly water ice) and micron-sized dust particles, a few kilometres in diameter¹¹. Because of their small size neither gravitational accretion energy nor internal radiogenic heating is sufficient to differentiate the nuclei¹², whose large formation distance from the sun and long-term storage in the Oort cloud precludes any subsequent possibility of extreme heating.

Interplanetary dust particles recovered by high-flying aircraft, which are presumed to have a cometary origin, seem to be agglomerations of interstellar dust grains and show no evidence of melting or differentiation¹³. Although comets may be a source for the very primitive carbonaceous chondrites, even those meteorites may be too processed to come from such a primitive source.

On the other hand, the differentiated meteorite types are believed to come from large parent bodies where gravitational and radiogenic heating is sufficient to have caused either partial or complete melting, or more modest-sized bodies where the heating has only been sufficient to drive off the more volatile elements and compounds and to cause partial melting of individual chondrite grains. Reflectance spectroscopy studies of meteorites and asteroids have shown many common analogues between the two, both in the main belt and among the known Earth-crossing asteroids (L. A. McFadden, M. J. Gaffey and T. B. McCord, personal communication). Thus, the data on impactor compositions at six of the seven periodic craters in the sample are more consistent with an asteroidal source for the impactors than with a cometary one.

Additionally, although no specific impact crater has been identified with the K-T extinction, Palme has found that the siderophile signature in the boundary clay "does not match the siderophile element pattern of unfractionated meteorites (chondrites)"⁹. An iron asteroid impactor has been suggested but is not a certainty.

It is difficult to imagine a dynamic mechanism that would supply asteroidal impactors to the Earth with a 26-28-Myr period. The typical dynamic half-life for bodies placed in Earth-crossing orbits is ~30 Myr. Impacts from any periodic injection of asteroids into Earth-crossing orbits would be scattered randomly in time over that period. What would be needed is a mechanism that both injected bodies into Earth-crossing orbits periodically and then destroyed them relatively quickly (in 1-5 Myr) if they were not swept up by the Earth.

If there is no dynamic mechanism for supplying asteroidal impactors with the suggested (or any) period, then one may

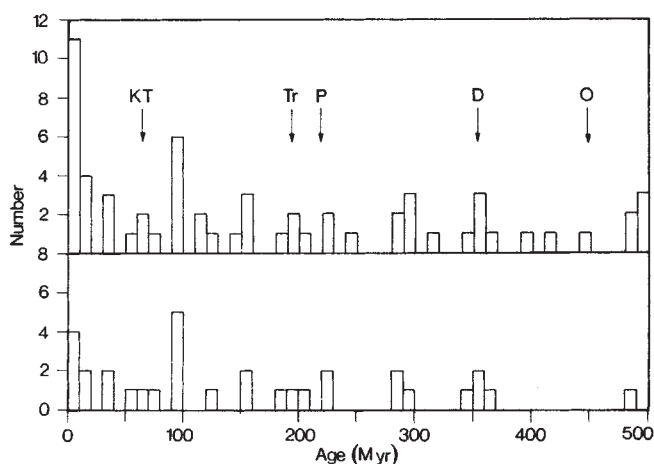


Fig. 1 Distribution of ages for dated craters on the Earth. Upper histogram is for all dated craters; lower histogram is for those with diameter ≥ 10 km. Arrows indicate the approximate times of the five major biological extinction events in the past 500 Myr.

suspect that the periodicity in the cratering and extinction records is not real but is a statistical fluke. Additionally, one may speculate that the extinction mechanism is not always impact related but also results from some other terrestrial or extraterrestrial event or process, such as the passage of the Solar System through a giant molecular cloud or close to a supernova explosion or an unusual period of terrestrial volcanism.

Another question is, if there are no periodic comet showers from the inner Oort cloud, is there still evidence of random showers as suggested by Hills¹⁴? The distribution of ages of known impact craters on the Earth over the past 500 Myr is shown in Fig. 1. The upper histogram includes 61 craters listed by Grieve⁶, without regard to the formal error on the age; the lower histogram includes only those 32 craters with diameters ≥ 10 km. Also shown are the approximate times of the five major extinctions in the past 500 Myr, as given by Sepkoski¹⁵.

There seem to be non-random clusterings of craters at certain times in the Earth's past, and in some cases these may be correlated with recognized extinction events. In particular, the five craters at ~360 Myr BP (four of them ≥ 10 km in diameter) are close in age to the late Devonian (D) extinction event; an iridium anomaly was recently identified at this boundary¹⁶. Another grouping of five smaller craters at ~490 Myr BP may be associated with one of the lesser extinction events identified by Sepkoski, the Cambrian biomere events. On the other hand, a cluster of five craters at ~290 Myr BP (three, ≥ 10 km) is not associated with any recognized extinction and the late Ordovician (O) extinction at ~440 Myr BP has no substantial clustering of impacts simultaneous with it. The record is less clear for the Permian (P), Triassic (Tr) and K-T events where there is a more continuous distribution of large and small craters in time.

The crater clusterings associated with the late Devonian and Cambrian biomere events may be evidence of random cometary showers in the past, caused by random stars passing through the inner Oort cloud. A random star might be expected to pass within 10^4 AU of the sun, the approximate transition distance between the inner and outer Oort cloud, every 80 Myr¹⁷. Of the 10 craters possibly associated with these two events, only the 5-km crater at Saaksjarvi, Sweden has an identified impactor type, a stony-iron. The lack of identified impactors is largely attributable to a failure to search for them. Higher velocity impacts, however, tend to leave a smaller fraction of the meteoritic material in the impact melts⁹, making the impactor type harder to identify. Cometary impacts are expected to be at average velocities of 29 km s^{-1} for short-period comets and 57 km s^{-1} for long-period comets, considerably greater than the $15\text{--}20 \text{ km s}^{-1}$ expected for asteroids¹⁸. On the other hand, the

crater clusterings may only be the result of random small-number statistics; more detailed analysis is clearly required.

It seems that there is not a simple one-to-one relationship between major asteroid and/or comet impacts, siderophile anomalies and biological extinction events on the Earth. Although impacts may be occasionally, or even often, the major triggering event, it is still possible that there are other types of events that may also trigger extinctions. Similarly, the secondary mechanism that actually causes the extinction, for example, long-term darkness, planetary cooling or atmospheric pollution, may vary from one extinction to another depending on the exact state of the planet and its biota at the time of the event. These questions remain open, awaiting new and better data for their resolution.

I thank Bonnie Buratti, George Wetherill, Richard Grieve, Frank Kyte and Mark Bailey for useful discussions and comments. This work was supported by the NASA Planetary Geosciences Program and was performed at the Jet Propulsion Laboratory under contract with NASA.

Received 2 November 1984; accepted 30 January 1985.

1. Raup, D. M. & Sepkoski, J. J. Jr *Proc. natn. Acad. Sci. U.S.A.* **81**, 805-805 (1984).
2. Alvarez, W. & Muller, R. A. *Nature* **308**, 718-720 (1984).
3. Whitmire, D. P. & Jackson, A. A. Jr *Nature* **308**, 713-715 (1984).
4. Davis, M., Hut, P. & Muller, R. A. *Nature* **308**, 715-717 (1984).
5. Rampino, M. R. & Stothers, R. B. *Nature* **308**, 709-712 (1984).
6. Grieve, R. A. F. *Geol. Soc. Am. spec. Pap.* **190**, 25-37 (1982).
7. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095-1108 (1980).
8. Bohor, B. F., Foord, E. E., Modreski, P. J. & Triplehorn, D. M. *Science* **224**, 867-869 (1984).
9. Palme, H. *Geol. Soc. Am. spec. Pap.* **190**, 223-233 (1982).
10. Weissman, P. R. in *Protostars and Planets II* (ed. Black D. C.) (University of Arizona Press, in the press).
11. Donn, B. & Rahe, J. in *Comets* (ed. Wilkening, L. L.) 203-226 (University of Arizona Press, 1982).
12. Lewis, J. S. *Icarus* **15**, 174-185 (1971).
13. Fraundorf, P., Brownlee, D. E. & Walker, R. M. in *Comets* (ed. Wilkening, L. L.) 383-409 (University of Arizona Press 1982).
14. Hills, J. G. *Astr. J.* **86**, 1730-1740 (1981).
15. Sepkoski, J. J. Jr *Geol. Soc. Am. spec. Pap.* **190**, 283-289 (1982).
16. Playford, P. E., McLaren, D. J., Orth, C. J., Gilmore, J. S. & Goodfellow, W. D. *Science* **226**, 437-439 (1984).
17. Weissman, P. R. *Nature* **288**, 242-243 (1980).
18. Weissman, P. R. *Geol. Soc. Am. spec. Pap.* **190**, 15-24 (1982).

Timescale of natural annealing in radioactive minerals affects retardation of radiation-damage-induced leaching

Yehuda Eyal*† & Robert L. Fleischer*

* General Electric Research and Development Center, Schenectady, New York 12301, USA

† Department of Nuclear Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel

There is concern that if solidified nuclear waste is exposed to ground water, accelerated dissolution of radioactive waste nuclides will ensue because of the accumulation of radiation damage created by α decay. We report here the results of radioisotope measurements on ancient uranium- and thorium-rich minerals and find that 'self-annealing' strikingly reduces the rate of dissolution. That short-lived ^{228}Th is leached, whereas long-lived ^{230}Th and ^{234}U are not, sets a limit on the timescale of self-annealing; given a simple assumption, we estimate an annealing time of 15 kyr. Two further conclusions emerge. First, even though the overall structure of the minerals remains crystalline (inferred from X-ray diffraction), there must be localized submicroscopic loss of order because preferential leaching takes place. Second, the inference that damage fades and slows the leaching rate indicates that minerals in which such annealing occurs are better waste-storage media for minimizing loss of radionuclides by dissolution.

Safe storage of solid nuclear waste requires a matrix that will not be easily dissolved by ground water if a buried container is breached. Although waste-storage materials that have minute

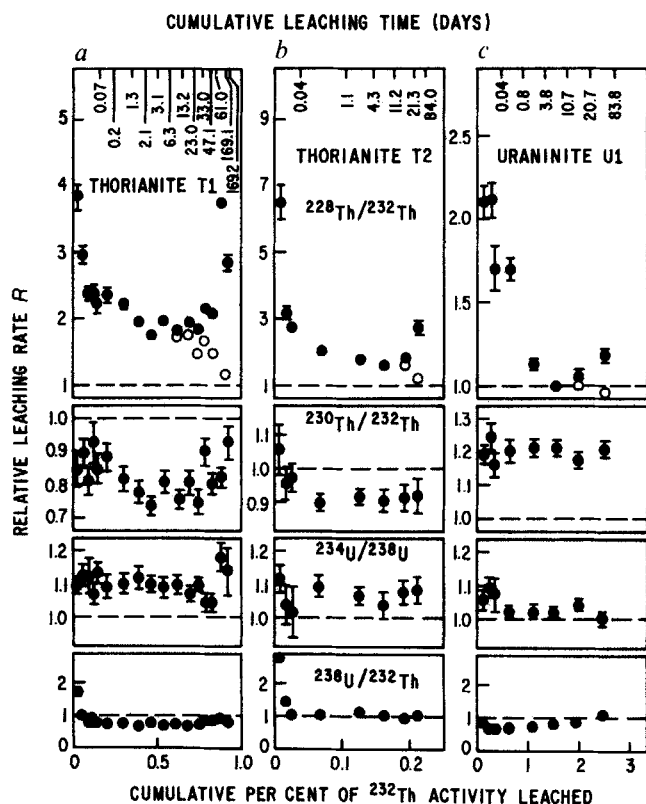


Fig. 1 Relative leaching rates R (solid circles) of uranium and thorium isotopes leached at 25 °C by a bicarbonate/carbonate solution from two thorianite specimens (a, b) and a uraninite (c). The value of R is the ratio of fractions of radioactivities leached. Values of $R(^{228}\text{Th}/^{232}\text{Th}) \gg 1$ show that the ^{228}Th lies in leachable radiation-damage sites, whereas values of $R(^{230}\text{Th}/^{232}\text{Th})$ and $R(^{234}\text{U}/^{238}\text{U})$ elevated by no more than 20% imply that the damage fades over considerably less time than the mean life of ^{234}U (3.5×10^5 yr). For the $^{228}\text{Th}/^{232}\text{Th}$ fractionation, the open circles give the lower limit of $R(^{228}\text{Th}/^{232}\text{Th})$, assuming the ancestor nuclide of ^{228}Th , ^{228}Ra , is completely retained in the solid phase.

leaching rates are available, it is possible¹ that the radioactivity that is to be isolated will itself cause radiation damage and thus bring about greatly accelerated chemical attack by ground water. The major source of damage is expected to result from α decay of heavy elements, which produces heavy residual nuclei that recoil a distance of ~ 200 Å, displacing $\sim 1,500$ atoms².

The evidence for accelerated dissolution by such damage comes both from laboratory experiments and field observations. Individual α -recoiling nuclei can be implanted in a crystal or glass and then released by exposure to water^{3,4}, implying that the damaged region is dissolved to where the implanted atoms come to rest and that doses of implanted low-energy ions sufficiently intense that the clumps of displaced atoms overlap, produce bulk dissolution to the depth of the implantation⁵. Ground water commonly leaches ^{234}U , in preference to ^{238}U , out of rocks that it permeates⁶, which is widely (but not universally) regarded as a result of the ^{234}U being located within α -recoil tracks⁷ produced by decay of ^{238}U . The ^{238}U atoms are at the start of the decay chain and are therefore not located in radiation-damaged material.

The preferential leaching in nature of ^{234}U (mean life 3.5×10^5 yr or half-life 2.4×10^5 yr) provides a test of the stability of α -recoil-produced radiation damage over times that are vastly greater than those attainable in the laboratory and in conditions that are relevant to underground storage of nuclear wastes. For minerals where the radiation damage is stable at ambient temperatures, the long-term accelerated ^{234}U leaching will be troublesome for disposal. However, in minerals where chemical attack is not accelerated, we have suitable candidates for waste

storage media⁸ which could well be useful in other high-radiation environments such as those at fission or fusion reactors.

In contrast to long-lived ^{234}U from primary ^{238}U , the thorium decay product of primary ^{232}Th , ^{228}Th , has a mean life of 2.7 yr. The latter nuclide is formed by decay of the α -recoil atom ^{228}Ra (mean life, 8.3 yr). Therefore, the leaching of ^{228}Th will display the effects of radiation damage that is always relatively fresh. A second radiogenic thorium isotope is available for comparison, namely, the long-lived ^{230}Th (mean life, 1.1×10^5 yr) from the uranium decay chain. We have observed profound differences between the behaviour of the $^{228}\text{Th}/^{232}\text{Th}$ pair and the pairs $^{230}\text{Th}/^{232}\text{Th}$ and $^{234}\text{U}/^{238}\text{U}$.

The stability of a material against radiation damage is closely correlated with its propensity to become amorphous (metamict) during irradiation⁹. Understanding what controls the ease of disordering is a basic question in solid-state physics¹⁰. For example, Ewing *et al.*¹¹ have shown how the leachability of zircon increases with the natural internal irradiation dose and Clinard *et al.*¹² have examined ^{238}Pu -doped zirconolite in which damage accumulates rapidly enough for laboratory experiments.

We report here leaching tests on two types of ancient radioactive minerals, thorianite and uraninite, composed primarily of highly-crystalline thorium-rich and uranium-rich ThO_2/UO_2 solid solutions, respectively. Structurally, these are the natural analogues of fluorite-type actinide dioxides, which are important nuclear-waste materials. Included are two thorianite specimens from Sri Lanka (samples T1 and T2) and a uraninite (sample U1) from Wilberforce, Ontario, Canada. Analysis by electron microprobe (EMP) showed that uranium, thorium and lead were uniform within the major phase of each mineral studied ($< 10\%$ variation) and that only small-volume fractions of accessory phases were present. The latter was confirmed by the absence of unexplained X-ray diffraction spots. The ThO_2 and UO_2 contents were determined by radiochemical analyses (procedures described in refs 13, 14). Lead isotope analyses with an ion microprobe, together with elemental analyses by the EMP, gave radiometric ages according to the U/Th/Pb methods. These data are given in Table 1, which also lists the number of times the average atom has been displaced by α -recoil and α -particle processes. The ages agree with those previously measured for other samples from the same locale¹⁵. Note that, if the displaced atoms remained displaced over all of time, the material would be thoroughly and uniformly damaged and therefore would be uniformly leached. But we find that this is not the case.

We subjected the finely pulverized (mean particle diameter of the order of 1 μm , maximum particle size of $\sim 100 \mu\text{m}$) mineral samples to a series of successive leaching treatments in a bicarbonate/carbonate solution (0.1 M NaHCO_3 /0.1 M Na_2CO_3) and then determined the relative leaching rates of ^{238}U , ^{234}U , ^{232}Th , ^{230}Th and ^{228}Th (see refs 13, 14). The enhanced dissolution, shown in Fig. 1, of ^{228}Th relative to its parent ^{232}Th , clearly indicates the presence of localized damage by α -recoil atoms. However, the much smaller isotope fractionation, on dissolution, between ^{234}U and its parent ^{238}U and between ^{230}Th and ^{232}Th suggests that almost complete annealing of the damage occurs in $< 10^5$ yr. Note that the initial reduction in the value of $R(^{228}\text{Th}/^{232}\text{Th})$ as dissolution proceeds may be attributed to the rapid depletion of ^{228}Th in the layer of the mineral that is in contact with the solution. However, further exposure of the mineral to the solution causes an increase in the value of $R(^{228}\text{Th}/^{232}\text{Th})$. As recently observed for the mineral monazite^{13,14}, the effect has been caused by retention of ^{228}Ra in the solid phase.

The presence of residual atomic-displacement damage in the minerals is further suggested by two consequences of annealing the samples at 810 °C for ~ 7 h; the leaching rates are lower and the lattice parameter decreases by 0.2–0.5% (full details are given elsewhere¹⁶).

The leaching results allow a damage storage-time to be derived. The emerging physical picture is of minerals in which each atom over the age of the mineral has been displaced from its lattice site many times; yet the mineral remains crystalline.

Table 1 Compositions, U/Th/Pb ages and radiation damage in thorianite and uraninite specimens

Mineral	ThO ₂ (wt %)*	UO ₂ (wt %)*	Age (Myr)†	Dpa‡
Thorianite				
T1	47.6	31.6	550	155
T2	51.9	26.9	550	141
Uraninite				
U1	14.8	61.3	1,000	450

* Uncertainties ~2%.

† Uncertainties ~7%. For specimens T1 and T2 a composite age is given.

‡ Estimated displacements per atom, assuming 1,500 atomic displacements per α decay in the uranium, actinium and thorium series.

The inescapable conclusion is that the damage has been healing itself. We now consider the consequences of a steady-state process of damage and recovery, in which severe localized damage from α -recoil events survives for some time before vanishing into a background of more dispersed damage. This mode generates vital annealing-time information. For all three mineral samples, T1, T2 and U1, the value of $R(^{228}\text{Th}/^{232}\text{Th})$ was found to be initially much greater than unity (3.8, 6.4 and 2.1, respectively, from Fig. 1), whereas the values of $R(^{234}\text{U}/^{238}\text{U})$ were 1.11, 1.12 and 1.08, respectively. As the ^{228}Th is short lived and ^{234}U is long lived, the simplest interpretation of these results is that self-annealing heals the radiation damage in a time between ~11 yr (the combined mean lives of ^{228}Ra and ^{228}Th) and $\sim 3.5 \times 10^5$ yr. Assuming that in the absence of annealing the $^{234}\text{U}/^{238}\text{U}$ fractionation would have been as high as the observed $^{228}\text{Th}/^{232}\text{Th}$ fractionation, then the annealing time is $\tau_{234} [R(^{234}\text{U}/^{238}\text{U}) - 1] / [R(^{228}\text{Th}/^{232}\text{Th}) - 1]$, where τ_{234} is the mean life of ^{234}U (3.5×10^5 yr) and R is the initial relative isotope leaching rate. (The derivation is given in more detail elsewhere¹⁶.) We find average annealing times¹⁶ of ~15 kyr (14, 7.8 and 25 kyr for the three samples). A time of 18 kyr is implied by the monazite data of Eyal^{13,14}, who, however, quoted only an upper limit of 10^5 yr.

Two conclusions emerge. First, even highly non-metamict minerals, as usually inferred from X-ray diffraction, may be subjected to radiation damage by α -recoil atoms. The localized submicroscopic variation in atomic order in these minerals may cause preferential leaching and, consequently, radioactive disequilibrium, as is often observed in nature. Second, the method of isotope analysis used here should allow the damage storage time of other minerals to be measured and improved waste-storage materials to be identified.

We thank R. B. Bolon, C. L. Chalek, R. P. Goehner, W. Katz, N. Lewis, D. W. Marsh, D. B. Sorensen and L. G. Turner for experimental assistance. Most of this work was done while Y.E. was a Visiting Fellow at the General Electric Research and Development Center.

Received 6 August 1984; accepted 16 January 1985.

1. Burns, W. G., Hughes, A. E., Marples, J. A. C., Nelson, R. S. & Stoneham, A. M. *Nature* **295**, 130-132 (1982).
2. Matzke, H. *Rad. Effects* **64**, 3-33 (1982).
3. Fleischer, R. L. *Science* **207**, 979-981 (1980).
4. Fleischer, R. L. *Geochim. cosmochim. Acta* **46**, 2191-2201 (1982).
5. Dran, J. C., Maurette, M. & Petit, J. C. *Science* **209**, 1518-1520 (1980).
6. Szabo, B. J. in *8th INQUA Conf.*, 941-949 (International Union for Quaternary Research, Brussels, 1969).
7. Huang, W. H. & Walker, R. M. *Science* **155**, 1103-1106 (1967).
8. Roy, R. & Vance, E. R. *J. Mater. Sci.* **16**, 1187-1190 (1981).
9. Mitchell, R. S. *Miner. Rec.* **4**, 177-182 (pt 1), 214-223 (pt 2) (1973).
10. Naguib, H. M. & Kelley, R. *Rad. Effects* **25**, 1-12 (1975).
11. Ewing, R. D., Haaker, R. F. & Lutze, W. in *Scientific Basis for Radioactive Waste Management V* (ed. Lutze, W.) 389-397 (Elsevier, Amsterdam, 1982).
12. Clinard, F. W. *et al. J. nucl. Matter* **105**, 248-256 (1982).
13. Eyal, Y. & Kaufman, A. *Nucl. Tech.* **58**, 77-83 (1982).
14. Eyal, Y. in *Scientific Basis for Radioactive Waste Management V* (ed. Lutze, W.) 399-408 (Elsevier, Amsterdam, 1982).
15. Faul, H. *Nucl. Geol.* (Wiley, New York, 1954).
16. Eyal, Y. & Fleischer, R. L. *Geochim. cosmochim. Acta* **49**, 5 (1985).

A thermo-mechanical model of continental lithosphere

Sean D. Willett*, David S. Chapman* & Horst J. Neugebauer†

* Department of Geology and Geophysics, University of Utah, Salt Lake City, Utah 84112-1183, USA

† Institut für Geophysik, Technische Universität Clausthal, Clausthal-Zellerfeld 3392, FRG

An understanding of the mechanical response of the continental lithosphere to vertical surface loading is crucial when considering geodynamic processes such as intraplate basin formation, post-glacial rebound or mountain building. Previous models of the mechanical behaviour of the lithosphere, which are more successful in oceanic than in continental regions, have used specific isotherms to delineate an effectively elastic zone of the lithosphere and result in mathematically convenient descriptions of mechanical behaviour. Heat flow and lithosphere flexure data for continents, however, suggest that this relationship is more complex, which can be demonstrated by numerical modelling of the loading of a heterogeneous visco-elastic lithosphere with temperature-dependent viscosity. Viscous stress relaxation in the lower and hotter section of the lithosphere results in a region of stress concentration dependent on the thermal state of the lithosphere, its structure and the time since loading. Here we present initial numerical models, based on a relatively simple steady-state thermal model constrained by present-day heat flow, which explain observations of continental flexure except in regions which have experienced a long or complex thermal history since loading.

As a first approximation, both oceanic and continental lithosphere loading have been modelled as a flexed elastic plate overlying an inviscid fluid substratum¹⁻³. But laboratory studies of the deformation of rocks⁴⁻⁶ and detailed modelling of the above phenomena⁷⁻¹⁰ suggest the actual rheology of the lithosphere must be considerably more complex. Based on studies of the deformation of olivine, Goetze and Evans⁷ proposed a rheology in which the uppermost lithosphere is weakened by brittle failure and stresses in the lower lithosphere are relaxed by a temperature-dependent non-linear dislocation flow mechanism, leaving most recoverable strain in the region between. They further suggested that the temperature-dependent flow could be approximated by a plastic failure criterion. This allows a time-independent calculation of stress distribution and, more specifically, predicts an observable property of apparent elastic lithosphere thickness. This yield-stress envelope model was applied with particular success to oceanic lithosphere flexure^{9,11,12} and assisted in explaining the apparent correlation between elastic thickness and depth to the 450 °C isotherm as calculated from the cooling plate model associated with sea-floor spreading^{2,13}.

Recently, Karner *et al.*¹⁴ applied this thermo-elastic model to continental lithosphere flexure. Their analysis of flexural rigidity of continental lithosphere in terms of thermal age of the lithosphere at the time of loading led them to conclude the following: (1) flexural rigidities for continental lithosphere are compatible with oceanic results and are explained universally by a simple thermo-elastic model; (2) the long-term thermal behaviour of continental lithosphere is governed by a simple cooling plate model with a 200-250-km plate thickness; (3) the effective elastic thickness of continental lithosphere coincides with the depth to the 450 °C isotherm at the time of loading and does not change with time. Although we find support in heat-flow data for the basic contention of Karner *et al.*¹⁴ that lithosphere mechanical behaviour is controlled largely by its thermal state, the precise relationship between thermal state and mechanical behaviour may be more complex than their conclusions suggest and we offer here an alternative explanation for continental flexure observations.

Table 1 is a compilation of continental flexure observations and estimates of present surface heat flow at each site. The latter values are generally averages of many heat-flow determinations with effects such as anomalous radiogenic heat production of surface rocks, or possible hydrological effects weighted so that the average heat flow is a reliable indicator of lithospheric thermal state. We make two observations from these data. First, the present thermal state of continental lithosphere at flexure sites is not a simple function of its age (compare column 6 with columns 3 and 4 in Table 1). This observation is consistent with results obtained from much larger heat-flow data sets not restricted to flexure sites¹⁵⁻¹⁷ and should be considered in judging the general applicability of a thermal evolution model. Second, apparent elastic thicknesses for continental flexure sites do not correspond to the 450 °C-isotherm position. In fact, by using present-day heat flow to compute the present thermal state, we see that elastic thicknesses correspond to depths for a variety of isotherms from 300 °C to >900 °C (Fig. 1). The thermal regime has not necessarily been constant since loading. But as a plastic yield-stress model responds to the warmest thermal state following loading, the data (Fig. 1) represent lower limits along the surface heat-flow axis. The observation of very high temperatures or of a wide range of temperatures controlling elastic thickness is inconsistent with the oceanic thermo-mechanical model.

Is there a rational explanation for the apparent success of the plastic yield-stress approximation in explaining oceanic lithosphere flexure but its lack of success for the continents? Courtney and Beaumont¹⁸ have demonstrated that the rapid cooling of oceanic plates effectively locks in the stress field and prohibits significant stress relaxation. In continental lithosphere subjected to a longer and more complex thermal history, we believe this total damping of the long-term relaxation may not be taking place. However, this seeming contradiction can be examined by obtaining complete time-dependent solutions to the more general visco-elastic problem.

To obtain a time-dependent solution of lithosphere flexure with a non-linear thermally activated rheology, we must resort to numerical methods. The model we have developed consists of a heterogeneous Maxwell visco-elastic lithosphere overlying an inviscid asthenosphere. The viscous strain rates $\dot{\epsilon}$ in the lithosphere are determined by a power law ($n=3$ in stress σ) which is exponentially dependent on absolute temperature T through an activation energy E^* , that is, $\dot{\epsilon} = A\sigma^n \exp(-E^*/RT)$, where A , n and E^* are properties of the material and R is the gas constant. For the mantle we have used the experimentally determined flow parameter for olivine given by Goetze⁵ ($A = 7 \times 10^4 \text{ MPa}^{-3} \text{ s}^{-1}$, $E^* = 0.53 \text{ MJ mol}^{-1}$). Para-

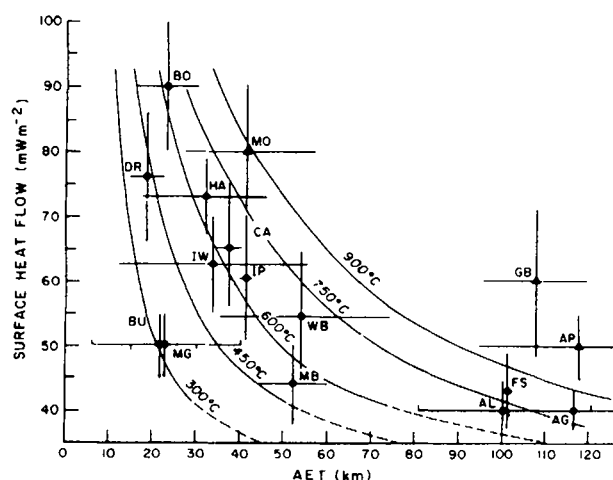


Fig. 1 Surface heat flow at selected flexure sites versus AET. Data and symbols from Table 1. Solid lines indicate depths to isotherms at a given heat flow for a steady-state thermal model.

meters for crustal rocks are difficult to determine because of the scarcity of data, but seem to have consistently lower activation energies. We used values of $A = 7 \times 10^{-6} \text{ MPa}^{-3} \text{ s}^{-1}$ and $E^* = 0.23 \text{ MJ mol}^{-1}$. The thickness of the lithosphere is determined by its thermal state; its base is taken to be approximately the position where the geotherm intersects the 1,300 °C adiabat. A vertical uniformly-distributed surface load of width equal to the lithosphere thickness and infinite length is applied to the upper surface. This load is supported by buoyancy forces resulting from flow in the mechanical asthenosphere. The mechanical strength of the lithosphere determines the redistribution of these opposing forces. We have used a finite element method to solve for the time-dependent displacements and stresses within the visco-elastic body incorporating an algorithm and code from Owen and Hinton¹⁹. (Details of this modelling are given in Willett *et al.*²⁰.)

Salient features of the flexure of a continental lithosphere with this rheology are shown in Fig. 2a, b. These two examples are chosen to represent end-member thermal states for continental lithosphere. Fig. 2a, labelled 'cold lithosphere', represents deformation of a lithosphere with a vertical temperature distribution corresponding to a steady-state geotherm with surface heat flow of 40 mW m^{-2} ; Fig. 2b, labelled 'hot lithosphere', has a 90 mW m^{-2} or active-rift geotherm²¹. The displacements in both cold and hot lithosphere show characteristics of visco-

Table 1 Continental lithosphere age, flexure and thermal characteristics*

Site	Symbol	Basement age at time of loading (Myr BP)	Age of load (Myr BP)	AET (km)	Present heat flow (mW m^{-2})	Heat-flow reference
Lake Agassiz	AG	1,600-1,800	0.007-0.014	101-132	40 ± 3	24
Lake Algonquin	AL	1,200-1,500	0.0011-0.0013	80-120	40 ± 5	24
Fennoscandia	FS	2,000-2,800	0.012	101	43 ± 7	25
Caribou Mts	CA	1,600-1,800	1-10	34-40†	65 ± 10	26
Interior Plains	IP	1,600-2,500	1-10	41†	60 ± 10	27
Ganges Basin	GB	1,000-2,000	10-50	95-120	60 ± 12	28, 29
Idaho-Wyoming Thrust	IW	75-275	50-150	12-55	62 ± 8	30
Michigan Basin	MB	800-1,100	370-460	44-60	44 ± 6	30
Boothia Uplift	BU	1,000-1,400	460-550	6-40†	50 ± 5	27
Midcontinent Gravity High	MG	100-400	1,100	15-30	50 ± 5	30
Lake Hamilton	HA	200	0.009-0.019	18-46	73 ± 6	31
Lake Bonneville	BO	200	0.015-0.020	16†	90 ± 10	32
North Great Dividing Range	DR	0-100	400-500	15-22	76 ± 10	33
Appalachian Foreland Basin	AP	560-1,000	330-390	95-140	50 ± 5	30
Molasse Basin	MO	230-320	20-30	26-56	80 ± 10	34
Williston Basin	WB	2,000-2,400	110-450	34-75‡	55 ± 10	35

* Modified from ref. 23; † data from ref. 1; ‡ data from ref. 36.

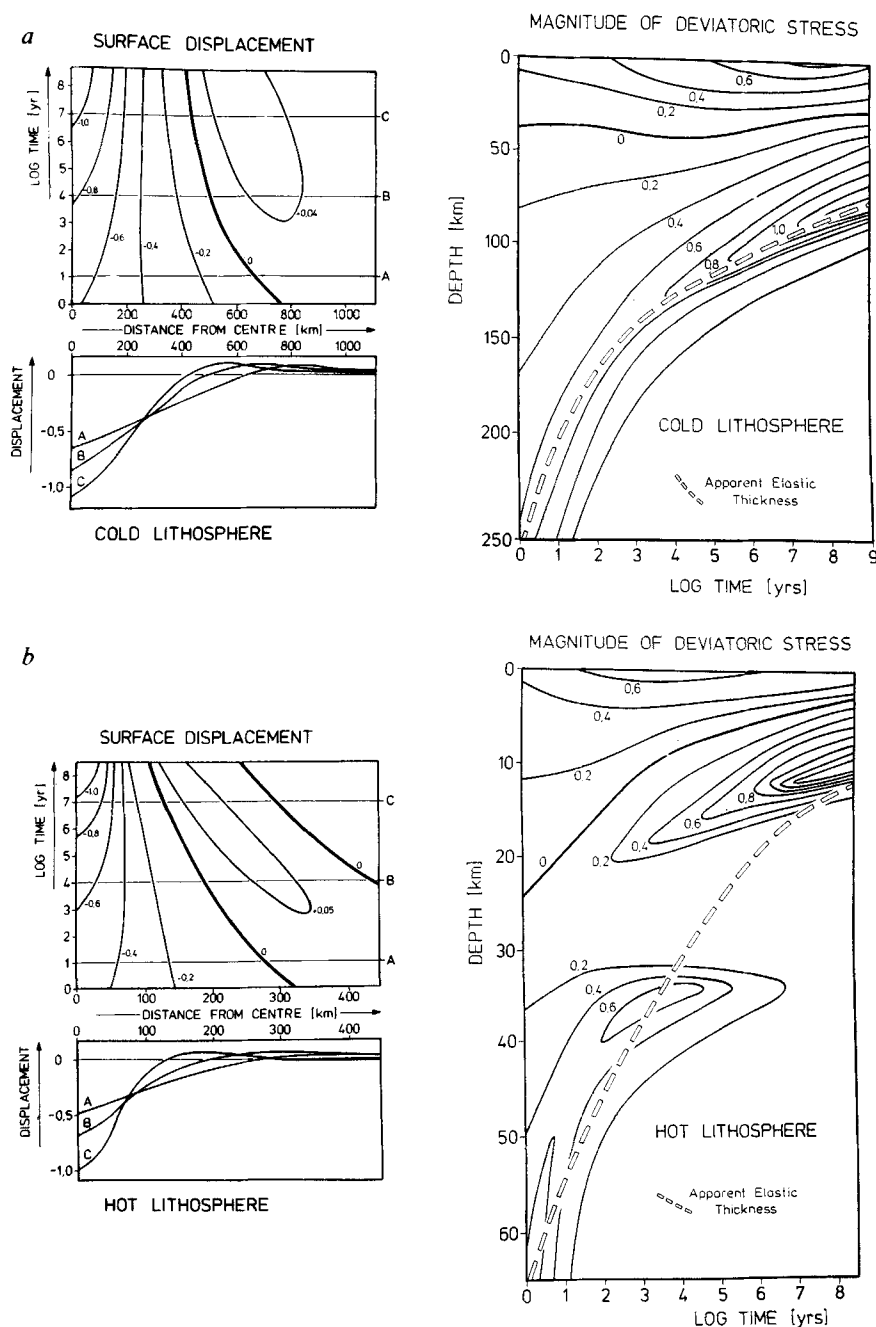


Fig. 2 Flexure properties for two end-member thermal states of visco-elastic continental lithospheres. In *a* and *b* the upper left figure shows surface vertical displacements mapped as a continuous function of time; lower left figure gives the original surface (that is, basement) profile at three distinct times. Right figures show deviatoric stresses on a plane directly below the centre of the load and the temporal evolution of this stress field following instantaneous loading at time zero. Dashed line gives apparent elastic thickness as calculated from the wavelength of the surface displacements. The 'cold lithosphere' has a vertical temperature distribution corresponding to a geotherm with a surface heat flow of 40 mW m^{-2} ; displacements are normalized to 1,200 m; stresses are normalized to 50 MPa. The 'hot lithosphere' has a temperature distribution corresponding to a geotherm with a surface heat flow of 90 mW m^{-2} ; displacements are normalized to 1,600 m; stresses are normalized to 100 MPa.

elastic plate flexure: a central depression, deepest at the centre, which deepens and narrows with increasing time; and the development of a region of vertical uplift, a peripheral bulge, which in time migrates towards the basin.

The stress field also shows the basic elements of flexural response. At any given time it includes a plane of zero deviatoric stress separating regimes of horizontal extension and compression. The magnitude of the stress reaches a maximum at the surface or base of the mechanical lithosphere. Immediately after the onset of loading, stresses at the base of the lithosphere begin to relax and a region of effectively zero deviatoric stress develops. As time progresses this relaxed region grows upward, thereby concentrating stress in the upper lithosphere, a phenomenon which has also been observed in other geodynamic models^{18,22}. The region of negligible deviatoric stress acts as the mechanical asthenosphere. Because of the strong temperature dependence of flow rates, the rate of upward concentration of stresses decreases with time, allowing the colder upper lithosphere to maintain stresses. This relaxation pattern is changed with an increase in heat flow and corresponding increase in temperature

as seen in the 'hot lithosphere' model. The increase in temperature activates flow in the lower crust and, consequently, stresses are relaxed quickly in the lower crust as well as in the lower lithosphere. This second zone of relaxation forces the upper 15–20 km to carry the entire load. The rheology of the crust controls lithosphere response in high temperature regimes.

The flexural wavelength of the surface displacements is used to calculate an apparent elastic thickness (AET) at each point in time. This AET has been superimposed on the stress fields of Fig. 2 and approximately corresponds to the maximum depth to which deviatoric stresses are maintained. Given the logarithmic scales in Fig. 2, it seems that little change in the stress and displacement fields takes place at longer times, an observation which supports the validity of an elastic-plastic model approximation. However, the long term value of elastic thickness is not a simple function of temperature and, until the processes determining this value are understood, we retain the complete solution.

For use in predicting flexural parameters, our model results are summarized in Fig. 3a. Each contour gives the apparent

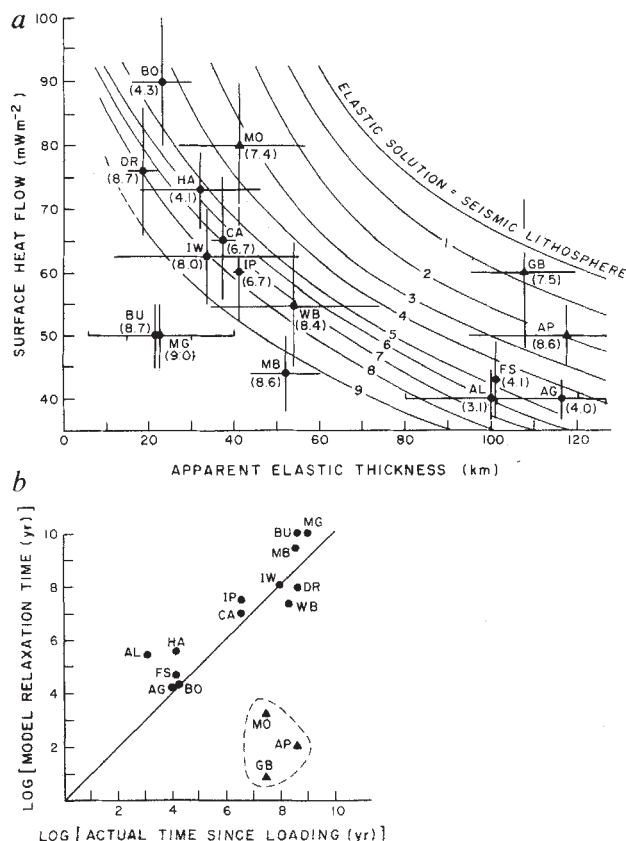


Fig. 3 *a*, Relaxation map relating apparent elastic thickness of a visco-elastic continental lithosphere to thermal state and loading time. Data and symbols from Table 1. Numbers in parentheses represent $\log_{10}$ (actual load duration in yr) for each site. Solid contour lines show the predicted weakening of the lithosphere with time after loading, caused by viscous stress relaxation in the lower and hotter part of the lithosphere; contour values are $\log_{10}$ (load duration in yr). Irregular contours result from lithosphere heterogeneity that affects the entire stress-relaxation process, and from numerical model interpolation. *b*, A comparison of the actual time since loading versus the visco-elastic model relaxation time necessary to produce the observed apparent elastic thickness. Model data are interpolated from relaxation map in *a*. Triangles are foreland basin sites in collision zones.

elastic thickness, as calculated from the visco-elastic model, for a given thermal state and a particular loading time. For example, a hot lithosphere characterized by surface heat flow of 90 mW m^{-2} (for example, Basin and Range) has an AET of 65 km for very short loading times (seismic lithosphere), which relaxes to $<30 \text{ km}$ in 10^4 yr (Lake Bonneville unloading) and

would further relax to an AET of 15 km were the load and the same thermal state to exist for 10^8 yr . In contrast, an intermediate 'warm' lithosphere characterized by surface heat flow of 60 mW m^{-2} starts out with a seismic lithosphere of 125 km, relaxes to an AET of 50 km in 10^6 yr after loading and to an AET of 38 km after 10^8 yr . We have also superimposed on this relaxation map (Fig. 3a) the flexure data from Fig. 1. Included with each data point is the age (duration) of the load. There is general agreement between actual load duration and model relaxation time necessary to reach the observed AET at a given thermal state, as shown by the near one-to-one correspondence in Fig. 3b, despite large uncertainties in both thermal and flexure data and the sensitivity of the relaxation map to model parameters. The fit in Fig. 3 is surprisingly good considering that the relaxation contours in Fig. 3a were calculated from models with general loading conditions and flow properties and a simple thermal model. Furthermore, as this relaxation model was constructed from steady-state thermal models, we would expect to fit only those data for which the thermal state has not changed appreciably since loading. For those data with long or complex thermal histories large errors in model fit could exist.

This is probably the case with the foreland basin data (MO, AP, GB), which depart dramatically from model predictions and misfit the model in Fig. 3. The lithosphere flexure in continental collision zones is the result of complex thermal and mechanical processes which apparently are not simulated by this model. It is less likely, but also possible, that heat-flow determinations surrounding foreland basins are systematically biased. The oldest loads (BU, MG, MB) have had a greater opportunity to experience transient thermal effects and so are more likely not to fit this steady-state model. This may be the case, but is obscured by the logarithmic scales in Fig. 3.

Our analysis of continental flexure data together with heat-flow data produces the following conclusions: (1) The application of oceanic lithosphere thermal and thermo-mechanical models to continental lithosphere does not seem to be supported by heat-flow and continental flexure data. In particular the oceanic lithosphere correlations between heat flow and age and between elastic thickness and a specific temperature range are not observed on continents. (2) Complete time-dependent solutions for visco-elastic lithospheric flexure with thermally activated flow properties can be obtained numerically, the principal free parameters being temperature and rheological properties. Generic visco-elastic mechanical models of a vertically heterogeneous lithosphere incorporating simple steady-state thermal models can explain most observed flexural parameters. Those data that are not well fit by the model have long or complex thermal or mechanical histories. An understanding of both is critical when considering the long-term behaviour of continental lithosphere.

We thank the Alexander von Humboldt Foundation (DSC) and the Deutscher Akademischer Austauschdienst (SDW) for support and the University of Utah Research Committee for financial assistance.

Received 31 July 1984; accepted 11 February 1985.

- Walcott, R. I. *J. geophys. Res.* **75**, 3941-3954 (1970).
- Caldwell, J. G. & Turcotte, D. L. *J. geophys. Res.* **84**, 7572-7576 (1979).
- Karner, G. D. & Watts, A. B. *J. geophys. Res.* **88**, 10449-10477 (1983).
- Ashby, M. F. & Verrall, R. A. *Phil. Trans. R. Soc. A* **288**, 59-95 (1977).
- Goetze, C. *Phil. Trans. R. Soc. A* **288**, 99-119 (1978).
- Kirby, S. *Rev. Geophys. Space Phys.* **21**, 1458-1487 (1983).
- Goetze, C. & Evans, B. *Geophys. J. R. astr. Soc.* **59**, 463-478 (1979).
- Watts, A. B., Bodine, J. H. & Steckler, M. S. *J. geophys. Res.* **85**, 6369-6376 (1980).
- McNutt, M. K. & Menard, H. W. *Geophys. J. R. astr. Soc.* **71**, 363-394 (1982).
- Quinlan, Garry M. & Beaumont, C. *Can. J. Earth Sci.* (in the press).
- Watts, A. B. & Talwani, M. *Geophys. J. R. astr. Soc.* **36**, 57-90 (1974).
- Bodine, J. H., Steckler, M. S. & Watts, A. B. *J. geophys. Res.* **86**, 3695-3707 (1981).
- Parsons, B. & Sclater, J. G. *J. geophys. Res.* **82**, 803-827 (1977).
- Karner, G. D., Steckler, M. S. & Thorne, J. A. *Nature* **304**, 250-253 (1983).
- Rao, R. U. M., Rao, G. V. & Reddy, G. K. *Earth planet. Sci. Lett.* **59**, 288-302 (1983).
- Chapman, D. S., Germann, D. & Neugebauer, H. J. *Terra Cognita* **3**, 95 (abstract) (1983).
- Morgan, P. & Sassi, J. H. *J. Geodynamics* **1**, 143-166 (1984).
- Courtney, R. C. & Beaumont, C. *Nature* **305**, 201-204 (1983).
- Owen, D. R. J. & Hinton, E. *Finite Elements in Plasticity*, (Pineridge, Swansea, 1980).
- Willett, S. D., Chapman, D. S. & Neugebauer, H. J. *Ann. Geophys.* (in the press).

- Pollack, H. N. & Chapman, D. S. *Tectonophysics* **38**, 279-296 (1977).
- Kusznir, N. S. & Park, R. G. *Nature* **299**, 540-542 (1982).
- Watts, A. B., Karner, G. D. & Steckler, M. S. *Phil. Trans. R. Soc. A* **305**, 249-281 (1982).
- Jessop, A. M. & Lewis, T. J. *Tectonophysics* **50**, 55-77 (1978).
- Cermak, V. in *Terrestrial Heat Flow in Europe* (eds Cermak, V. & Rybach, L.) 3-40 (Springer, Heidelberg 1979).
- Majorowicz, J. A. & Jessop, A. M. *Tectonophysics* **74**, 209-238 (1981).
- Jessop, A. M., Hobart, M. A. & Sclater, J. G. *Geotherm. Ser. 5*, (Geothermal Service of Canada, Earth Physics Branch, Ottawa, 1976).
- Rao, R. U. M., Rao, G. V. & Narain, H. *Earth planet. Sci. Lett.* **30**, 57-64 (1976).
- Rao, G. V. & Rao, R. U. M. *Earth planet. Sci. Lett.* **48**, 397-405 (1980).
- Sass, J. H. et al. in *McGraw-Hill/CINDAS Data Series on Material Properties Vol. II-2* (eds Touloukian, Y. S., Judd, W. R. & Roy, R. F.) 503-548 (McGraw-Hill, New York, 1981).
- Lewis, T. J. & Jessop, A. M. *Can. J. Earth Sci.* **18**, 366-375 (1981).
- Chapman, D. S. *Abstracts of papers presented at Interdisciplinary Symposia* **14**, 436 (XVII General Assembly of the IUGG, Canberra, Australia, 1979).
- Cull, J. P. & Denham, D. *Bur. Miner. Resources J. Aust. Geol. Geophys.* **4**(1), 1-13 (1979).
- Haelen, R. in *Atlas of Subsurface Temperatures in the European Community* (eds Haelen, R. et al.) (Th. Schaefer, Hannover, 1980).
- Gosnold, W. D. *Jr Bulletin No. 84-04-MMRRI-04*, 1-110 (Mining and Mineral Resources Research Institute, University of North Dakota, Grand Forks, 1984).
- Ahern, J. L. & Mrkwicka, S. R. *Tectonics* **3**, 79-102 (1984).

Lateral transport of Mn in the north-east Pacific Gyre oxygen minimum

J. H. Martin & G. A. Knauer

Moss Landing Marine Laboratories, Moss Landing, California 95039, USA

Accurate measurement of the concentration and fluxes of a wide range of dissolved and particulate materials in the water column is needed to quantify the global cycling of heat, water and chemical constituents¹. An important aspect of any Global Ocean Flux study is the lateral transport of materials at basin or global scales; to study lateral transport, useful tracers of water-mass movements must be identified. Here we report that Mn is rapidly removed ($-0.027 \mu\text{mol m}^{-3} \text{yr}^{-1}$) from the north-east Pacific Gyre dissolved-Mn maximum onto biogenic particles sinking through the oxygen minimum (500–1,100 m). To support removal, dissolved Mn must be supplied within the oxygen minimum at net advective horizontal transport rates of $\sim 0.6\text{--}0.9 \text{ cm s}^{-1}$. The high affinity of dissolved Mn for particles, together with the readily discernible differences in inshore/offshore dissolved distributions, indicate that Mn is a chemical tracer with great potential in future studies of large-scale circulation processes in the ocean¹.

We show dissolved and suspended particulate Mn distributions observed 830 km north of the Hawaiian Islands (28°N ; 155°W), together with temperature, salinity and dissolved oxygen data, in Fig. 1. Dissolved-Mn concentrations were relatively high (0.8 nmol kg^{-1}) in near surface waters, then decreased steadily until a minimum ($\sim 0.2 \text{ nmol kg}^{-1}$) was reached at depths of 200–450 m (probably originally subarctic surface waters). Dissolved Mn began increasing concomitantly with rapidly decreasing oxygen concentrations in the 450–700-m depth interval. After reaching a maximum of $0.54 \text{ nmol kg}^{-1}$ at 700 m, concentrations decreased to $0.34 \text{ nmol kg}^{-1}$ at 1,000 m. With the exception of two relatively high values observed at 1,090 m ($0.55 \text{ nmol kg}^{-1}$) and 1,420 m ($0.40 \text{ nmol kg}^{-1}$), amounts of dissolved Mn remained fairly constant ($0.24\text{--}0.29 \text{ nmol kg}^{-1}$) for the remainder of the 2,090-m water column sampled. These data were obtained using GoFlo-bottle sampling, Chelex-100 ion-exchange preconcentration and atomic absorption analytical procedures that are described in detail elsewhere².

In addition to the dissolved amounts, two fractions of suspended particulate Mn were also measured: weakly leachable (LE, that dissolved with 25% acetic acid)³; and that remaining after leaching (RE, the refractory portion, dissolved via Teflon bomb digestion)⁴. The latter is believed to be that locked into aluminosilicate

lattices and, as expected, amounts were extremely low ($0.2\text{--}5.9 \text{ pmol kg}^{-1}$) in these remote gyre waters far removed from continental land masses (Fig. 1). Particulate LE Mn concentrations were very low ($4\text{--}9 \text{ pmol kg}^{-1}$) in the upper 140 m of the water column (Fig. 1). More or less steadily increasing values were found from 140–600 m, followed by a minimum at 900 m, a secondary maximum at 1,090 m and steadily decreasing concentrations thereafter.

Because free-floating particle traps⁵ were also used during this study (deployed for 33 days), flux data are available to determine whether Mn was being regenerated from (decreasing fluxes with depth), or scavenged onto (increasing fluxes with depth), particles sinking through the oxygen minimum. The dramatic increases in Mn fluxes (Fig. 2) indicate, in view of the following considerations, that dissolved Mn was being removed onto particles. Increases in total flux did not result from the lateral transport of aluminosilicate phases because RE Mn fluxes remained uniformly low and constant ($2\text{--}3 \mu\text{mol m}^{-2} \text{yr}^{-1}$) throughout the 2,000-m water column (Fig. 2). Furthermore, the RE Mn fluxes were similar to atmospheric air/sea Mn fluxes ($3.3 \mu\text{mol m}^{-2} \text{yr}^{-1}$) estimated by Duce at Eniwetok⁶, and we assume that this was the primary source of the aluminosilicate (that is, RE) Mn in our traps.

Trapped exchangeable (TE, that dissolved in the particle trap cyclinders)^{2,7,8} Mn fluxes were also fairly constant with depth ($14\text{--}20 \mu\text{mol m}^{-2} \text{yr}^{-1}$), though they were an order of magnitude greater than the RE Mn fluxes. Clearly then, the fraction responsible for the dramatic increase in total Mn fluxes between 500 and 1,000 m was the LE fraction (Fig. 2). This flux increase could result from *in situ* scavenging of dissolved Mn or from the advective lateral transport of particles already rich in LE Mn. Comparisons of suspended particulate LE Mn and trapped particulate LE Mn concentrations (Fig. 1) indicate that the trapped particles are remarkably different from those in suspension (at least judging by their Mn content) for most of the water column; that is, Mn concentrations are not equivalent in the two classes of particles until depths of 900 m are reached. Thus, we conclude that the increase in flux does indeed result from the *in situ* removal of dissolved Mn onto biogenic particles sinking through the O_2 minimum/Mn maximum.

Assuming steady-state conditions, this removal of dissolved Mn must be balanced by input processes of some kind. To define the importance of various input mechanisms, we used a simple 500–1,100-m box model shown in Fig. 3. The model illustrates the following: (1) Mn removed from the box by particle scavenging at a rate of $-0.027 \mu\text{mol m}^{-3} \text{yr}^{-1}$ (500-m(LE+TE)Mn flux minus the 1,100-m(LE+TE) Mn flux divided by $1,100 - 500 \text{ m}$ = 600 m) would be completely removed in 16 yr because the average integrated Mn concentration in the box is only

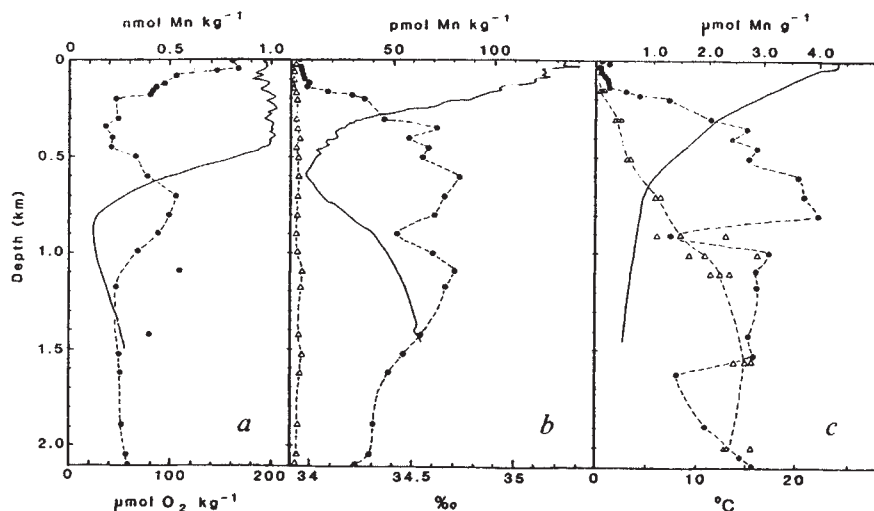


Fig. 1 Manganese and hydrographic measurements made 830 km north of Hawaii (28°N ; 155°W) in July/August of 1983: a, dissolved Mn (—●—) and oxygen (—); b, suspended particulate leachable (—●—) and refractory (—△—) Mn with salinity (—); c, suspended (—●—) and trapped (—△—) particulate Mn concentrations on a gram-dry-weight basis, together with temperature (—) data.

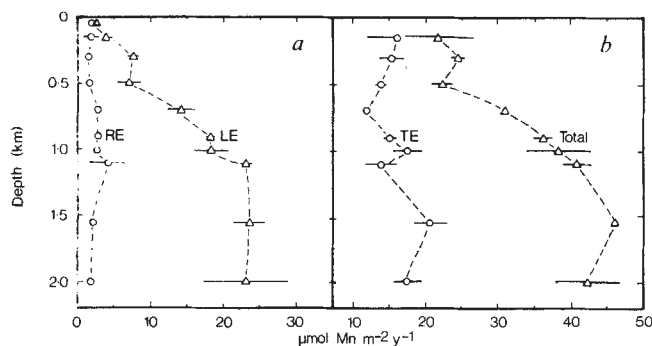


Fig. 2 Manganese fluxes measured with particle traps at 28°N, 155°W: *a*, particulate refractory (RE) Mn (—○—) and particulate leachable (LE) Mn (—△—); *b*, trapped exchangeable (TE, Mn dissolved in trap solutions) (—○—) and total Mn fluxes (—△—). Means and standard deviations ($n=3$) are shown for all depths except 700 m (LE, $n=2$, range is shown; TE and total, $n=1$). Standard deviations for the refractory portions were usually smaller than the symbols used; 50-m TE and total fluxes are not shown because of possible zooplankton swimmer⁵ contamination.

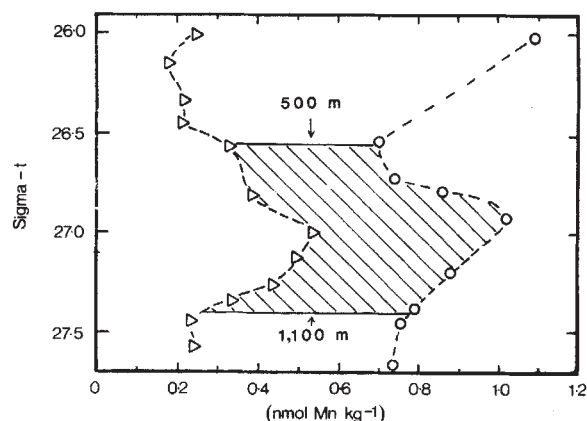


Fig. 4 Dissolved Mn versus density (σT) north of Hawaii (Δ 28°N, 155°W) and west of California ($\circ$ 36.5°N, 123°W). Hatched area represents Fig. 3 box model based on Hawaii depth-versus- σT relationship.

$0.42 \mu\text{mol m}^{-3}$. (2) The shape of the dissolved-Mn profile (Figs 1, 3) shows that vertical advective/diffusive processes cannot be re-supplying the Mn removed by scavenging. Dissolved Mn advecting upward into the box at 1,100 m is balanced by Mn advected upward out of the box at 500 m because the concentrations at 500 and 1,100 m are about the same and there is no reason to believe that vertical advection rates would be markedly different at these depths (for example, we use a velocity of 4 m yr^{-1} at both depths)⁹. (3) It is also apparent that diffusive mixing processes are removing Mn from, rather than supplying Mn to, the box; $\Delta\text{Mn}/\Delta z$ calculations from 400–600 m and 1,000–1,200 m result in an upward Mn flux of $5 \mu\text{mol m}^{-2} \text{ yr}^{-1}$ across the 500-m top of the box and a downward flux of $4 \mu\text{mol m}^{-2} \text{ yr}^{-1}$ across the 1,100-m bottom of the box when a global ocean average eddy diffusivity constant of $1.7 \text{ cm}^2 \text{ s}^{-1}$ is used¹⁰. (4) Adding up the total vertical fluxes into ($22 \mu\text{mol m}^{-2} \text{ yr}^{-1}$) and out of ($-47 \mu\text{mol m}^{-2} \text{ yr}^{-1}$) the box results in a difference of $25 \mu\text{mol m}^{-2} \text{ yr}^{-1}$ that must be supplied

by horizontal advective/diffusive processes to maintain steady-state conditions.

We believe that the source of the Mn being removed in the north-east Pacific Gyre is provided by Mn regeneration^{2,7,11,12} along continental margins. For example, dissolved-Mn profiles measured off California¹³ are remarkably similar in shape to those measured in this study, when they are compared with sigma- t (σT) data (see Fig. 4). Concentrations are also significantly higher inshore, which must be so if onshore/offshore lateral transport arguments are to be used.

We have no evidence that the Mn in the gyre indeed originated off the west coast of the United States. However, assuming that inshore Mn concentrations along other north Pacific continental margins are about the same as those we show in Fig. 4 and that the distance from California Station to the gyre station is representative (3,150 km), we can make first-order transport estimates assuming that $\Delta\text{Mn}/\Delta t = (\Delta\text{Mn}/\Delta x)u$, where $\Delta\text{Mn}/\Delta t$ is the combined average removal of dissolved Mn from the 500–1,100-m box by particulate scavenging and vertical mixing processes ($-25 \mu\text{mol m}^{-2} \text{ yr}^{-1} \div 600 \text{ m} = -0.042 \mu\text{mol m}^{-3} \text{ yr}^{-1}$); $\Delta\text{Mn}/\Delta x$ is the average difference between inshore and offshore dissolved-Mn concentrations in the oxygen minimum ($0.46 \mu\text{mol m}^{-3}$, from Fig. 4) divided by the distance ($3.15 \times 10^6 \text{ m}$) between the two locations; and u is the resulting net advective velocity in units of m yr^{-1} .

Using these estimates, the net advective velocity necessary to support Mn removal within the oxygen minimum off Hawaii is $290,000 \text{ m yr}^{-1}$ or 0.9 cm s^{-1} . Based on particulate scavenging rates alone (ignoring vertical mixing processes), the estimates become $180,000 \text{ m yr}^{-1}$ or 0.6 cm s^{-1} . Thus, the average length of time required to transport the dissolved Mn from California to Hawaii would be ~ 10 –20 yr.

Clearly, the net advective transport rates presented above are crude first estimates. We considered Mn transport only along the x and z axes, but dissolved Mn may be gained or lost by lateral mixing processes along the y axis as well. Nevertheless, with appropriate expansion of the water column and particle-trap sampling grid, and the simultaneous measurements of other trace elements that are removed in the open ocean (for example, J.H.M. and G.A.K., in preparation) or along continental margins (for example, Ag, Cu)¹⁴ as well as radionuclides (^{210}Pb , ^{231}Pa , ^{230}Th)^{15,16}, we believe that the use of these chemical tracers will result in a significant gain in the understanding of global ocean circulation processes.

Finally, we believe that the primary processes responsible for the excess Mn accumulating on the sea floor^{17,18} are the combination of continental weathering input, dissolved-Mn lateral transport and sinking biogenic particle scavenging output. Although

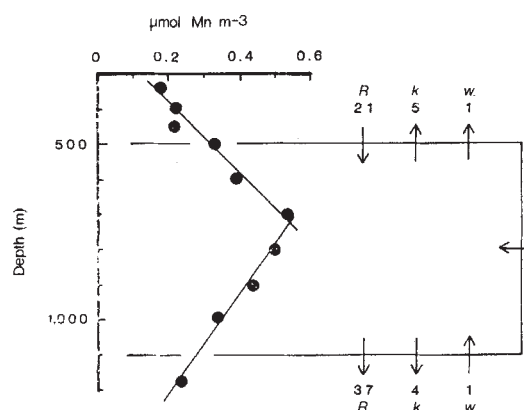


Fig. 3 A box model for the 500–1,100-m depth interval with fluxes (arrows) in units of $\mu\text{mol Mn m}^{-2} \text{ yr}^{-1}$. Dissolved Mn versus depth regressions shown on the left side yielded equations of $\mu\text{mol Mn m}^{-3}$: $= -0.18 + 0.0010 (m)$ for the upper line ($r = 0.9758$; $n = 6$) and $= 1.05 - 0.0007 (m)$ for the lower line ($r = 0.9890$; $n = 5$). Vertical diffusive fluxes (k) were estimated using these slopes and an eddy diffusivity coefficient of $5,400 \text{ m}^2 \text{ yr}^{-1}$ ($1.7 \text{ cm}^2 \text{ s}^{-1}$). Vertical advective fluxes (w) across 500 and 1,100-m boundaries were estimated by multiplying Mn concentrations by a vertical advective rate of 4 m yr^{-1} . Particulate associated (LE + TE) Mn flux data (R) are from Fig. 2. The lateral flux (u) was estimated from the difference (total fluxes out of box minus total fluxes into box).

we have presented these arguments before in descriptions of research performed at pronounced oxygen minimum^{2,8} and/or nearer the shore^{7,19}, final proof required the strong evidence provided here of at depth lateral transport and significant particulate scavenging of dissolved Mn at an open ocean location.

We thank M. Gordon, S. Fitzwater, C. Hunter and S. Tanner for their help with the field and laboratory work and W. Broenkow for the hydrographic data. This research was supported by grants from the NSF Marine Chemistry Programme (VERTEX OCE 82-16670 and OCE 82-16671).

Received 22 June 1984; accepted 29 January 1985.

1. Steele, J. H. *EOS* **65**, 684 (1984).
2. Martin, J. H. & Knauer, G. A. *Earth planet. Sci. Lett.* **67**, 35-47 (1984).
3. Chester, R. & Hughes, M. J. *Chem. Geol.* **2**, 249-262 (1967).
4. Eggemann, D. W. & Betzer, P. R. *Analyt. Chem.* **48**, 866-890 (1976).
5. Knauer, G. A., Martin, J. H. & Bruland, K. W. *Deep-Sea Res.* **26**, 97-108 (1979).
6. Buat-Menard, P. in *Air Sea Exchange of Gases & Particles* (eds Liss, P. S. & Stinn, W. G. N.) 455-532 (Reidel, Holland, 1983).
7. Martin, J. H. & Knauer, G. A. *Deep-Sea Res.* **30**, 411-423 (1983).
8. Martin, J. H. & Knauer, G. A. *J. mar. Res.* **40**, 1213-1225 (1982).
9. Broecker, W. S. *Chemical Oceanography* (Harcourt Brace Jovanovich, New York, 1974).
10. Li, V.-H., Peng, T. H., Broecker, W. S. & Ostlund, H. G. *EOS* **63**, 971-972 (1982).
11. Trefry, J. H. & Presley, B. J. *Geochim. cosmochim. Acta* **46**, 1715-1726 (1982).
12. Kremling, K. *Nature* **303**, 225-227 (1983).
13. Knauer, G. A., Martin, J. H. & Gordon, R. M. *Nature* **297**, 49-51 (1982).
14. Martin, J. H., Knauer, G. A. & Gordon, R. M. *Nature* **305**, 306-309 (1983).
15. Bacon, M. P., Spencer, D. W. & Brewer, P. G. *Earth planet. Sci. Lett.* **32**, 277-296 (1976).
16. Anderson, R. F., Bacon, M. P. & Brewer, P. G. *Earth planet. Sci. Lett.* **66**, 73-90 (1983).
17. Bender, M. L., Ku, T.-L. & Broecker, W. S. *Earth planet. Sci. Lett.* **8**, 143-148 (1970).
18. Bender, M. L., Klinkhammer, G. P. & Spencer, D. W. *Deep-Sea Res.* **24**, 799-812 (1977).
19. Martin, J. H. & Knauer, G. A. *Earth planet. Sci. Lett.* **51**, 266-274 (1980).

Sr isotope composition of sea water over the past 75 Myr

M. R. Palmer*† & H. Elderfield‡

* Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK

‡ Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ, UK

The $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of ancient seawater, as recorded in marine carbonates, is an important tracer of long-term variations in ocean chemistry¹⁻⁵. However, the Sr isotope balance of the oceans has been difficult to constrain; consequently, attempts to evaluate the temporal $^{87}\text{Sr}/^{86}\text{Sr}$ changes have been largely qualitative. To constrain the causes of these variations we have measured $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in carefully cleaned unrecrystallized foraminifera from DSDP sites 21 and 357. The data presented here have been quantitatively modelled taking advantage of recent advances in understanding of the Sr geochemical cycle. They suggest that whereas hydrothermal fluxes and carbonate recycling are of major importance in defining the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio, the major control over its variations through the Cenozoic has been changes in the isotope composition of Sr derived from the weathering of silicate rocks.

Previous studies have suggested that the major controls over variations in the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio are: (1) the proportions of oceanic Sr derived from crustal (mainly granitic) and mantle (continental basalts and hydrothermal input) sources; (2) the rate of transport of Sr from the continents to the oceans; and (3) recycling of Sr from marine carbonate sediments^{2,5-9}. Two factors require that these ideas and the data behind them be re-examined. First, an increased understanding of the magnitude of the hydrothermal processes at ocean ridges allows greater constraints to be placed on the oceanic Sr isotope balance¹⁰⁻¹³. Second, a recently-produced Sr isotope curve for seawater shows more fluctuations through time than were previously thought to be present¹.

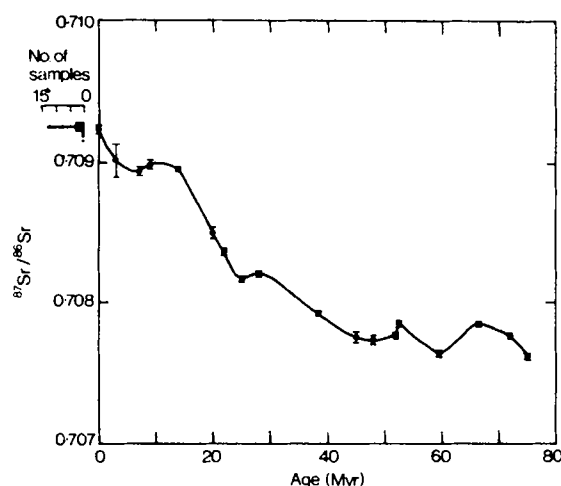


Fig. 1 Marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio since 75 Myr BP. Histogram is for 22 modern samples. National Bureau of Standards 987 SrCO_3 $^{87}\text{Sr}/^{86}\text{Sr} = 0.710275$, Eimer and Amend SrCO_3 $^{87}\text{Sr}/^{86}\text{Sr} = 0.708066$.

The results of this study are shown in Fig. 1 and Table 1. The form of the curve is in good agreement with the recently produced Burke curve¹. A further aspect of Fig. 1 is that it seems probable that the seawater Sr isotope curve can be constrained to a higher precision than is implied by the earlier studies through careful sample selection and cleaning. This considerably strengthens the potential of the curve for stratigraphical correlation and absolute age determinations during periods showing the greatest rates of change of $^{87}\text{Sr}/^{86}\text{Sr}$ ratios ($4 \times 10^{-5} \text{ Myr}^{-1}$). Measurements of 24 Recent foraminifera samples (see histogram in Fig. 1) are identical to within 2.6×10^{-5} (2σ), indicating a notional resolution of $\pm 0.65 \text{ Myr}$ (the practical limit of resolution is about $\pm 1.5 \text{ Myr}$). Thus, temporal variations in marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratios can provide a high-resolution dating and palaeostratigraphical method for the Oligocene-Recent, with potential for other periods as well.

To examine the various processes that may control the marine $^{87}\text{Sr}/^{86}\text{Sr}$ curve, a model was constructed based on the marine geochemical cycle of Sr shown in Fig. 2. In the model, the major flux of dissolved Sr to the oceans is from river input, $\sim 75\%$ of which is derived from weathering of uplifted limestones with the remainder resulting from the weathering of silicate rocks^{5,14}. Recycling also takes place through pore-water Sr diffusing into bottom waters from carbonate-rich sediments; a smaller flux is associated with sediments depleted in carbonate¹⁵. Sr isotope exchange takes place between basalt and seawater during circulation of seawater through fast-spreading mid-ocean-ridge hydrothermal systems⁷⁻⁹; no significant Sr mass transfer has been observed during this process. Neither Sr isotope exchange nor mass transfer is observed during lower temperature ridge-flank hydrothermal circulation¹¹ and the only significant sink for Sr in the modern oceans is biogenic calcium carbonate.

The present-day Sr fluxes are given in Fig. 2. Two values are given for the hydrothermal flux because of uncertainties over the proportions both of advective and convective heat flow at mid-ocean ridges and of advective heat loss associated with high-temperature ridge crest versus low-temperature ridge flank hydrothermal systems^{13,16-18}. One value is derived from linking chemical data in hydrothermal vents with global heat and ^3He loss from the ocean crust¹⁰⁻¹⁹; the second value acknowledges the conclusions from recent geophysical and geochemical work (mass balance of alkalis and alkaline earths, Li in basalts, Ge/Si ratios) that the hydrothermal flux at ridge axes is considerably lower than originally estimated^{12,13,20-23}. Although several studies have been made of the Sr isotope composition of river water^{4,5,24}, these have been limited to a few of the world's major rivers; a latest estimate from Canadian rivers²⁴, however, does suggest that the global average is likely to be ~ 0.711 . By solving

† Present address: Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA.

Table 1 $^{87}\text{Sr}/^{86}\text{Sr}$ and Sr/Ca ratios in foraminiferal calcite from DSDP sites 21 and 357

Site	Depth (m)	Age (Myr)	$^{87}\text{Sr}/^{86}\text{Sr} \pm 2\sigma$	Sr/Ca ($\times 10^3$)
357	1	0	0.709242 ± 27	1.35
357	11	3	0.709019 ± 121	1.27
357	29	7	0.708942 ± 28	1.15
357	39.5	9	0.708986 ± 20	1.19
357	50	14	0.708955 ± 14	1.19
357	105.5	20	0.708502 ± 40	1.37
357	138	22	0.708373 ± 15	1.26
357	194	25	0.708167 ± 16	1.24
357	210	28	0.708207 ± 27	1.23
257	258	38.5	0.707924 ± 14	1.32
357	287.5	45	0.707758 ± 37	1.38
357	380	48	0.707738 ± 29	1.42
357	444.5	52	0.707772 ± 18	1.17
357	467	52.5	0.707853 ± 18	0.99
357	476.5	59.5	0.707641 ± 22	1.12
21	92	66.5	0.707854 ± 12	1.18
21	108	72	0.707771 ± 16	1.08
21	128	75	0.707618 ± 17	1.36

the steady state equations (see Fig. 2) for the riverine Sr isotope composition, using the data in Fig. 2, we obtain $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of 0.7119 and 0.7102 (for the two different hydrothermal fluxes), both similar to the above value. However, the ratio required to balance the low hydrothermal flux yields a $^{87}\text{Sr}/^{86}\text{Sr}$ ratio for silicate weathering (0.7173) that is significantly closer to the average Sr isotope composition of aluminosilicate detritus in oceanic core top sediments (0.7166)²⁵ than does the ratio based on the high hydrothermal flux (0.7243).

Next, we considered temporal variations in the fluxes shown in Fig. 2. Oceanic crustal production rates are generally thought to have been higher 75 Myr BP than today^{26,27}. If we assume that the hydrothermal flux is directly proportional to seafloor spreading rate²⁸, then variations in the hydrothermal Sr flux can be calculated. Three estimates of seafloor spreading rates were considered here: rates that were 60% higher 65 Myr BP than today²⁹, rates that were a maximum of 25% higher 75 Myr BP than today³⁰ and an intermediate rate³¹ (corrected for subduction).

Because of the high concentration of Sr in marine carbonate, recycling of Sr from the weathering of uplifted limestones and diffusion of Sr from carbonate-rich sediment pore waters play an important part in the Sr marine geochemical cycle. From the mass half life of continental limestones¹⁴ and the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of ancient seawater, Brass⁵ calculated the average $^{87}\text{Sr}/^{86}\text{Sr}$ ratios derived from the weathering of uplifted limestones at intervals over the past 400 Myr. The $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of pore waters diffusing into overlying waters from carbonate-rich sediments is equivalent to that of 18-Myr BP seawater¹⁵. Thus, similarly to Brass, we calculated the average $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of this source against time.

The effects of carbonate recycling and varying hydrothermal fluxes on the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio since 75 Myr BP, obtained by these methods, are shown in Fig. 3a. It is clear that these two controls cannot fully explain the observed Sr isotope variations. Moreover, the consequence of assuming the most rapid variations in spreading rates²⁹ is that they produce unrealistically high global atmospheric CO_2 levels, temperatures and sea levels 75 Myr BP³² and, as noted above, recent evidence suggests that hydrothermal fluxes are lower than originally supposed. Hence, changes resulting from fluctuations in hydrothermal activity as well as from carbonate recycling appear to be minor. Thus, variations in the nature of the riverine Sr flux to the oceans must be a major control on temporal variations in the seawater $^{87}\text{Sr}/^{86}\text{Sr}$ ratio.

The lower $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in the past have been ascribed to lower river fluxes because of the decreased land area at that time^{7,14}. However, the rate of continental weathering is a com-

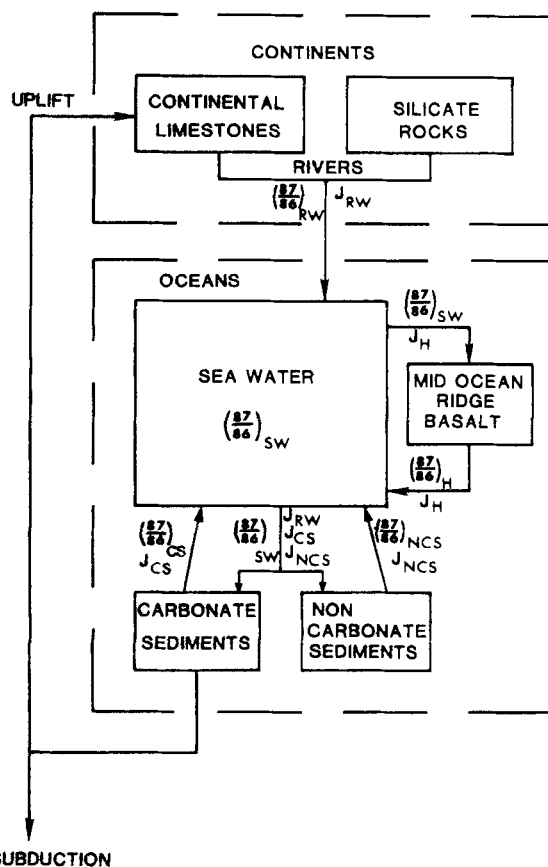


Fig. 2 Marine geochemical cycle of Sr. (J = flux; $(^{87}/^{86})$ = $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of flux.) At steady state: $[^{87}]_{\text{SW}} = ([^{87}]_{\text{RW}}J_{\text{RW}} + [^{87}]_{\text{H}}J_{\text{H}} + [^{87}]_{\text{CS}}J_{\text{CS}} + [^{87}]_{\text{NCS}}J_{\text{NCS}})/(J_{\text{RW}} + J_{\text{H}} + J_{\text{CS}} + J_{\text{NCS}})$; $[^{86}]_{\text{SW}} = ([^{86}]_{\text{RW}}J_{\text{RW}} + [^{86}]_{\text{H}}J_{\text{H}} + [^{86}]_{\text{CS}}J_{\text{CS}} + [^{86}]_{\text{NCS}}J_{\text{NCS}})/(J_{\text{RW}} + J_{\text{H}} + J_{\text{CS}} + J_{\text{NCS}})$; where SW = seawater, RW = river water, H = hydrothermal isotope exchange, CS = carbonate sediment pore waters, NCS = non-carbonate sediment pore waters. Magnitudes of present-day fluxes (mol yr^{-1}) are: $J_{\text{RW}} = 2.5 \times 10^{10}$ (ref. 37), $J_{\text{NCS}} = 4.0 \times 10^8$ (ref. 15), $J_{\text{CS}} = 3.0 \times 10^9$ (ref. 15), $J_{\text{H}} = 1.2 \times 10^{10}$ (ref. 10) = 0.38×10^{10} (ref. 11). $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of present-day fluxes are: $(^{87}/^{86})_{\text{NCS}} = 0.7064$ (ref. 15), $(^{87}/^{86})_{\text{CS}} = 0.7087$ (ref. 15), $(^{87}/^{86})_{\text{H}} = 0.7040$ (refs 8, 9), $(^{87}/^{86})_{\text{SW}} = 0.70924$ (this work). Sensitivity of model to uncertainties in above values; resulting uncertainty in calculated $(^{87}/^{86})_{\text{SW}}$ is 2×10^{-4} for uncertainty in J_{RW} and 1.4×10^{-4} for uncertainty in $(^{87}/^{86})_{\text{H}}$; all other variables produce negligible uncertainties in $(^{87}/^{86})_{\text{SW}}$ or are considered in the text.

plex function of climatic and geographical variables³³. Indeed, global CO_2 models suggest that the rate of weathering per unit area of continent was considerably higher 75 Myr BP than today because of the interplay of these factors³². It is likely, therefore, that the riverine Sr flux was at least equal to or higher than its present-day value. Thus, the low observed marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratios relative to the model values (Fig. 3) must reflect one or more of the following factors: a higher internal/external continental drainage ratio 75 Myr BP; decreases in the proportion of Sr derived from the weathering of limestones since 75 Myr BP; increases in the average $^{87}\text{Sr}/^{86}\text{Sr}$ ratio derived from the weathering of silicate rocks since 75 Myr BP.

The increase in land area since 75 Myr BP³⁴ requires extreme assumptions to be made for a concurrent fall in the continental internal/external drainage ratio. The Sr derived from limestone weathering would have to have been 10 times higher than that now to account for the smallest model deviation from the observed marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratios at 75 Myr BP (assuming a constant Sr flux magnitude and ratio from silicate weathering). Clearly, this mechanism alone cannot realistically account for the model deviations. These two arguments strongly suggest that

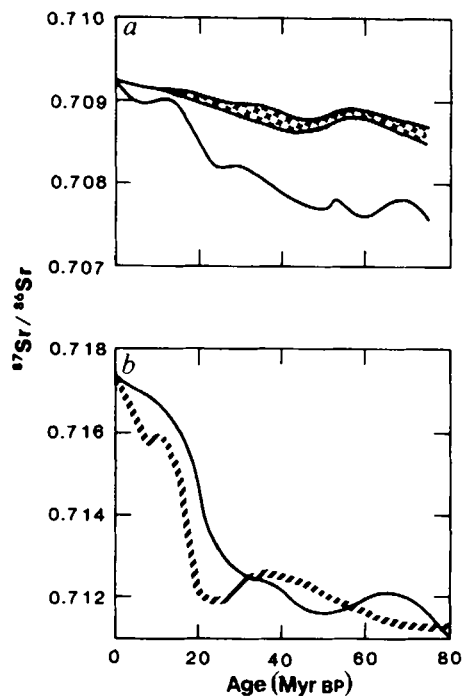


Fig. 3 *a*, Variations in the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio due to changes in rates of hydrothermal activity and carbonate recycling. Results using an intermediate rate of change in J_H (ref. 31) are shown (high and low rates of change in J_H (refs 29, 30) show similar patterns). The shaded area gives the range in $^{87}\text{Sr}/^{86}\text{Sr}$ ratios calculated from high and low values of J_H (refs 10, 11) and the bottom curve shows the measured values of $^{87}\text{Sr}/^{86}\text{Sr}$ for sea water. The model curves were calculated at 10-Myr intervals. *b*, Variations in the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio derived from the weathering of silicate rocks: the shaded area shows qualitative variations in glacial intensity (after Armstrong⁶).

changes in the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio derived from the weathering of silicate rocks must play a dominant role in varying the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio.

Figure 3*b* shows the changes in silicate $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (at a constant riverine flux magnitude) required to produce the observed marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratios, assuming a low hydrothermal flux and smaller changes in seafloor spreading rates. The most notable feature of the curves in Fig. 3*b* is the sharp rise in $^{87}\text{Sr}/^{86}\text{Sr}$ ratios from 25 Myr BP to the present, with relatively constant Sr isotope ratios before this time. This feature correlates well with two important variables in geological history. First, the similarity to the qualitative curve of glacial activity over the past 75 Myr BP is striking, providing some circumstantial support for Armstrong's⁶ hypothesis that increased glacial action has affected the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio by both exposing and increasing the weathering rates of ancient shield areas; these terrains have high $^{87}\text{Sr}/^{86}\text{Sr}$ ratios³⁵. Second, there is overwhelming evidence to suggest that during the late Cretaceous and early Cenozoic total worldwide volcanism was greater than at any subsequent time (see, for example, ref. 36). The past 30 Myr are marked by the Alpine and Himalayan orogenies. These widespread folding events resulted in the uplifting and accelerated erosion of rocks characterized by high $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Thus, the rise in the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio derived from silicate weathering since 25 Myr BP may reflect a recovery from the previous era when weathering of young basaltic volcanics was more dominant.

We conclude, therefore, that whilst hydrothermal fluxes and carbonate recycling are of great importance in determining the absolute value of the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio at any one instant, the major control over the variations in the ratio over the past 75 Myr has been changes in the isotope composition of Sr derived from silicate rock weathering.

We thank colleagues at Cambridge, Leeds and Yale, those associated with DSDP programme and J. Veizer for his review. This research was supported by NERC studentship GT4/80/GS/45 (M.R.P.) and grant GR3/4444 (H.E.). Cambridge Earth Sciences Series, contribution no. 541.

Received 10 September 1984; accepted 15 February 1985.

- Burke, W. H. *et al. Geology* **10**, 516-519 (1982).
- Peterman, Z. E., Hedge, C. E. & Tourtelot, H. A. *Geochim. cosmochim. Acta* **34**, 105-120 (1970).
- Dasch, E. J. & Biscaye, P. E. *Earth planet. Sci. Lett.* **11**, 201-204 (1971).
- Veizer, J. & Compston, W. *Geochim. cosmochim. Acta* **38**, 1461-1484 (1974).
- Brass, G. W. *Geochim. cosmochim. Acta* **40**, 721-730 (1976).
- Armstrong, R. L. *Nature phys. Sci.* **230**, 132-133 (1971).
- Spooner, E. T. C. *Earth planet. Sci. Lett.* **31**, 167-174 (1976).
- Albarede, F., Michard, A., Minster, J. F. & Michard, G. *Earth planet. Sci. Lett.* **55**, 229-236 (1981).
- Elderfield, H. & Greaves, M. J. *Geochim. cosmochim. Acta* **45**, 2201-2212 (1981).
- Edmond, J. M. *et al. Earth planet. Sci. Lett.* **46**, 1-18 (1979).
- Bender, M. L., Hess, J., Schilling, J. G. & Waggoner, D. G. *Trans. Am. geophys. Un.* **63**, 1014 (1982).
- Mottl, M. J. *Bull. geol. Soc. Am.* **94**, 161-180 (1983).
- Sleep, N. H. *Hydrothermal Processes at Seafloor Spreading Centers* (Plenum, New York, 1983).
- Garrels, R. M. & MacKenzie, F. T. *Evolution of Sedimentary Rocks* (Norton, New York, 1971).
- Elderfield, H. & Gieskes, J. M. *Nature* **300**, 493-497 (1982).
- Wolery, T. J. & Sleep, N. H. *J. Geol.* **84**, 249-275 (1970).
- Anderson, R. N., Langseth, M. G. & Sclater, J. G. *J. geophys. Res.* **86**, 3391-3409 (1977).
- Sclater, J. G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **86**, 11535-11552 (1981).
- Jenkins, W. J., Edmond, J. M. & Corliss, J. B. *Nature* **272**, 156-158 (1978).
- Hart, S. R. & Staudigel, H. *Earth planet. Sci. Lett.* **58**, 202-212 (1982).
- Seyfried, W. E., Janecky, D. R. & Mottl, M. J. *Geochim. cosmochim. Acta* **48**, 557-569 (1984).
- Sleep, N. H., Morton, J. L., Burns, L. E. & Wolery, T. J. *Hydrothermal Processes at Seafloor Spreading Centers* (Plenum, New York, 1983).
- Stoffyn-Egli, P. & Mackenzie, F. T. *Geochim. cosmochim. Acta* **48**, 859-872 (1984).
- Wadleigh, M. A., Veizer, J. & Brooks, C. *Geochim. cosmochim. Acta* (in the press).
- Dasch, E. J. *Geochim. cosmochim. Acta* **33**, 1521-1552 (1969).
- Pitman, W. C. & Talwani, M. *Bull. geol. Soc. Am.* **83**, 619-646 (1972).
- Hays, J. D. & Pitman, W. C. *Nature* **246**, 18-82 (1973).
- Delaney, M. L. thesis, Massachusetts Inst. Technology (1983).
- Pitman, W. C. *Bull. geol. Soc. Am.* **89**, 1389-1403 (1978).
- Parsons, B. *J. geophys. Res.* **87**, 289-302 (1982).
- Southam, J. R. & Hay, W. W. *J. Geophys. Res.* **82**, 3825-3842 (1977).
- Berner, R. A., Lasaga, A. C. & Garrels, R. M. *Am. J. Sci.* **283**, 641-683 (1983).
- Holland, H. D. *The Chemistry of the Atmosphere and the Oceans* (Wiley, New York, 1978).
- Vail, P. R., Mitchum, R. M. & Thompson, S. *Am. Ass. Petrol. Geol.* **26**, 83-97 (1978).
- McCulloch, M. T. & Wasserburg, G. J. *Science* **200**, 1003-1011 (1978).
- Ronov, A. B. *Geochim. Int.* **13**, 172-195 (1976).

Pine wood origin for pitch from the *Mary Rose*

R. P. Evershed, K. Jerman & G. Eglinton*

Organic Geochemistry Unit, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

The raising of the wreck of King Henry VIII's flagship (AD 1509-45), the *Mary Rose*, from the bed of the Solent was a unique event, notable as a significant technical achievement in naval architectural history and as a source of well-preserved Tudor relics. Preservation is attributed to the anaerobic environment prevailing within the sediment during much of the period of burial and, in some cases, to the large quantities of pitch that permeated the ship and many of the relics. Although the function of many relics is usually easily recognized, their means of construction and the nature and origin of the materials used in their manufacture are often much less obvious. We report here the chemical analysis of six samples of pitch from the *Mary Rose*. Computerized gas chromatography/mass spectrometry (CGC/MS) and infrared spectroscopy have been used to 'fingerprint' the pitches and compare them with modern-day tars and pitches derived from wood, coal, peat and petroleum. Diterpenoid hydrocarbons, methyl dehydroabietate and dehydroabietic acid were found in similar proportions to 'Stockholm tar' (good-quality pine tar obtained by the destructive distillation of *Pinus sylvestris*), so providing conclusive evidence for the derivation of the *Mary Rose* pitches from pine wood.

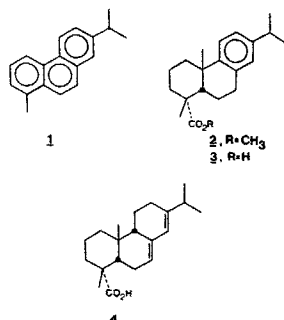
*To whom correspondence should be addressed.

Until recently, chemical archaeological investigations have been concerned largely with the inorganic components of, for example, building materials, pottery, coins and weapons¹. This apparent bias towards inorganic materials probably arises from their greater likelihood of survival over archaeological time compared with organic substances. Many inorganic materials remain largely unchanged after tens or even hundreds of years of weathering, whereas organic-based plant and animal products are rapidly decomposed unless protected from the degradative effects of moisture, oxygen and sunlight.

Historically, the use of tars and pitches as protective hydrophobic coatings is well established. However, to our knowledge no detailed chemical investigations have been performed to determine the origins or means of manufacture of such materials. Related organic chemical archaeological studies include GC analysis of animal fats from peat bogs and Eskimo dwelling sites^{2,3}, plant and animal residues from caves and pottery⁴, a ¹³C-NMR investigation of non-metallic seals⁵, an infrared study of tar from a thirteenth century Norwegian shipwreck⁶ and investigations of natural resins of art and archaeology, including GC and GC/MS studies^{7,8}.

Six samples of pitch and pitch-impregnated materials from the *Mary Rose* were selected for analysis. These included pitch from barrels found in the hold, 'luting' from the main deck, caulking from between the timbers of the outer frame of the ship, tarred anchor cable and a tarry solid (probably intrusive) found in a box of longbows. Table 1 lists these samples and their reference numbers.

GC and GC/MS analysis of the components of the *Mary Rose* pitches eluting in hexane, toluene and dichloromethane (DCM), and those obtained by esterification of the acid fraction, revealed in every case relatively simple mixtures of compounds. The hexane-eluting components of the pitch from the barrels produced the most complex patterns (Fig. 1a, b), with eight components comprising >90% of the total volatile components (calculated by integration of the reconstructed ion current (RIC) peak areas). The mass spectra of components 1-8 in Fig. 1 displayed molecular ions (M^+) at m/z 256 ($C_{19}H_{28}$), 254 ($C_{19}H_{26}$), 256 ($C_{19}H_{28}$), 252 ($C_{19}H_{24}$), 238 ($C_{18}H_{22}$), 220 ($C_{17}H_{16}$), 234 ($C_{18}H_{18}$) and 248 ($C_{19}H_{20}$), respectively. Comparison of these with reference spectra⁹ indicated that the com-



pounds were alkyl-substituted tricyclic diterpenoid hydrocarbons, the major component (peak 7 in Fig. 1) being retene (compound 1).

Further GC/MS analyses revealed the presence of methyl dehydroabietate (compound 2, m/z 314, $C_{21}H_{30}O_2$) as the major (>80%) ester component of the LC fractions from the *Mary Rose* samples. This latter compound was also the major constituent (>90%) of the methylated acid fraction, indicating that the principal acid is dehydroabietic acid (compound 3).

Diterpenoid compounds of this type have long been known to be major constituents of conifer resins^{10,11}, hence these CGC/MS data provide conclusive evidence that the *Mary Rose* pitches were produced from pine trees. As such compounds are not appreciably observed in peat and coal tars or petroleum bitumens, these can be discounted as potential sources. GC and CGC/MS of fractionated Stockholm tar produced a very similar distribution of hydrocarbons (Fig. 1c) to those of the barrels of pitch from the *Mary Rose*. Methyl dehydroabietate was also the

Table 1 *Mary Rose* tars and pitches and other materials

Sample	Appearance and odour	Origin
'Tarred' rope (MR AC)	20 strands 0.5-20 cm long, impregnated with tar	Anchor cable
'Pitch' (82 S 1298)	Shiny black conchoidally fractured fragments with strong 'tarry' odour	Barrel in the hold
'Luting' (MR 82 S1274/12)	Small pieces of damp tarred wood with a strong 'tarry' odour	Main deck
Caulking (MR 82 S126)	Rope-like material composed of hair fibres impregnated with tar	Between the timbers forming the outer frame
'Tar' (82 S1296/7)	Thick black tarry material with 'tarry' odour	Barrel in the hold
Solid (81 S300)	Soft dark-brown solid, smelling strongly 'tarry'	Inside chest/box of longbows (probably intrusive)

Bracketed codes were used by the *Mary Rose* Trust to document the samples. For simplicity the term 'pitch' has been used in the text but the terms 'tar' and 'pitch' are not synonymous. Tar is the initial pyrolysate, which may then be thickened to pitch by boiling to drive off volatiles and effect a degree of further polymerization. Pitches of different softening points are obtainable in this way, but the present study does not attempt to distinguish between them or tars.

major ester constituent and dehydroabietic acid again the major component of the acid fraction.

The remarkable similarities between the composition of the *Mary Rose* samples and 'Stockholm tar' clearly indicate that the sixteenth century and present-day tars and pitches are of the same generic type. These chemical data are in agreement with the historical and geographical evidence for the probable sources of the *Mary Rose* pitch¹².

Although the hydrocarbon distributions of the homogeneous pitch contained in the barrels are obviously similar to modern 'Stockholm tar', those obtained from the pitches already in their intended site of application (for example, 'luting', caulking and anchor rope) are somewhat different and variable¹². Nevertheless, they do comprise only the characteristic diterpenoid hydrocarbons referred to above. The differences in distribution may reflect varying degrees of microbial alteration, weathering or water washing during the ship's 36-yr active life and subsequent 437-yr submergence and burial. Alternatively, they may reflect differing means of preparation before use.

It is notable that none of the abietic acid (compound 4) commonly observed in pine resins^{10,11} was detected in either the *Mary Rose* samples or the 'Stockholm tar'. Presumably, the abietic acid present in the wood is thermally dehydrogenated and, to a lesser extent, decarboxylated during the retorting process, to produce the mixture of diterpenoid hydrocarbons, methyl dehydroabietate and dehydroabietic acid described above.

Today Stockholm tar or similar tars are imported into the United Kingdom from Russia, Eastern Europe, Scandinavia, China, America and India. At the time of the *Mary Rose*, it is known that trade developed between Russia and England, and amongst the first cargos were tar and pitch^{13,14}. The similarity in chemical composition between the *Mary Rose* samples and the modern Russian Stockholm tar is consistent with the *Mary Rose* pitch having originated from one such cargo, but a local source of manufacture cannot be ruled out.

We have recently examined further pitch samples from an excavation in York (early medieval) and an Etruscan shipwreck (600 BC) discovered off the Island of Giglio, Italy. Both exhibited infrared spectra similar to those of Stockholm tar and *Mary Rose* pitch. CGC/MS analyses have also shown them to

contain abundant diterpenoid compounds. Clearly, pine wood derived tars and pitches have been in use as maritime waterproofing agents for well over 2,000 yr.

This investigation demonstrates the power of GC and GC/MS in the analysis of the organic components of archaeological materials and illustrates the scope of the problems awaiting the application of modern analytical methodology.

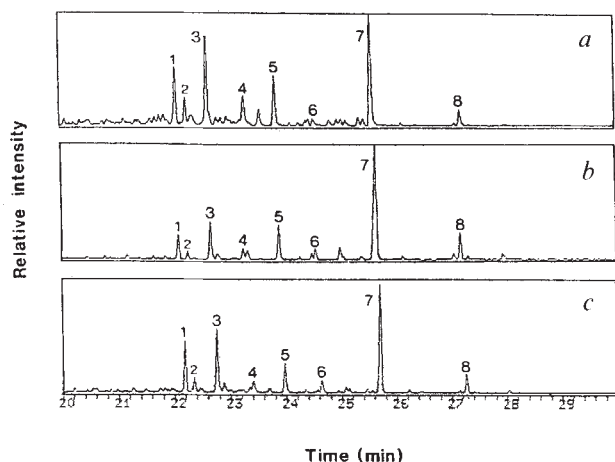


Fig. 1 Partial gas chromatograms showing the hydrocarbon fractions of: *a*, Mary Rose tar (82 S 1296/7); *b*, Mary Rose pitch (82 S 1298); *c*, present-day 'Stockholm tar' obtained by the destructive distillation of pine wood of the genus *Pinus*. Peak assignments of the alkyl-substituted tricyclic diterpenoid hydrocarbons were by GC/MS using a Finnigan 9610 GC linked to a Finnigan 4000 quadrupole MS. Data acquisition and processing were performed using a Finnigan INCOS 2300 data system. Peak identities are: 1 = 19-norabieta-8,11,13-triene (M^+ 256, $C_{19}H_{28}$); 2 = unknown (M^+ 254, $C_{19}H_{26}$); 3 = 18-norabieta-8,11,13-triene (M^+ 256, $C_{19}H_{28}$); 4 = simonellite (? M^+ 252, $C_{19}H_{24}$); 5 = 1,2,3,4-tetrahydrotene (M^+ 238, $C_{18}H_{22}$); 6 = mixture, predominantly ethylmethylphenanthrene (M^+ 220, $C_{17}H_{16}$); 7 = retene (compound 1) (M^+ 234, $C_{18}H_{18}$); 8 = unknown higher homologue of retene (M^+ 248, $C_{19}H_{20}$). Where the Mary Rose pitches were neither adhered to, nor incorporated into, any other materials, no extraction was necessary. In the cases of the caulking, 'luting' and rope samples, their associated pitch required extraction before analysis, achieved by placing the sample in an elution tube plugged with pre-washed cotton wool and washing with DCM (~200 ml). Extraction was continued until no further coloration appeared in the solvent washings. Removal of the DCM by rotary evaporation rendered the pitch extract free from particulate matter. Preliminary examination of the Mary Rose pitches and extracts was performed by infrared spectrophotometry as either melted films or as solutions in DCM. The infrared spectra of all the Mary Rose samples were similar, displaying absorptions assignable to carboxylic acid groups and aliphatic and aromatic moieties, and are very different from those recorded for petroleum bitumens and coal tars, but show some similarity to that obtained from a peat tar. The most striking similarities were observed between the Mary Rose samples and commercial 'Stockholm tar', whose spectrum is comparable in almost every respect (cm^{-1} and absorption intensity) to those of the Mary Rose pitches. Further comparisons required detailed molecular analysis by GC and GC/MS. In preparation for this, aliquots (~1 g) of the various tars and pitches were fractionated by liquid chromatography (LC, alumina, Al_2O_3) using gradient elution (100 ml each of *n*-hexane, toluene, DCM, ethyl acetate and methanol). Isolation of the acidic components of a Mary Rose tar (82 S 1298) and 'Stockholm tar' was achieved by shaking aliquots (0.5 g) of tar dissolved in DCM (5 ml) with aqueous potassium hydroxide (KOH, 5%, 4 × 5 ml). Neutralization (5 M HCl) of the aqueous extract, and extraction (DCM, 3 × 20 ml) produced the acids, which were converted to their corresponding methyl ester derivatives using ethereal diazomethane. Analyses were performed on a Carlo Erba 4160 GC, using on-column injection. The column was 25 m × 0.3 mm i.d. OV-1 coated flexible silica capillary column (Chromapak). Hydrogen was the carrier at 100 $cm s^{-1}$ and the GC oven temperature was programmed over 50–300 °C at 6 °C min^{-1} . The GC runs were acquired on a VG Analytical Minichrom data system¹⁵.

We thank the Mary Rose Trust for provision of samples and especially Mr I. Oxley and Dr M. Rule for valuable discussions. We also thank Mr M. Bound and the World Ship Trust for pitch from the Giglio wreck (courtesy of the Superintendency of Archaeology for Tuscany) and York Archaeological Trust for supplying the early medieval caulking; Dr W. Ladner and colleagues of the Coal Research Establishment for reference samples of coal and peat tars; Mr M. I. Danischewsky of the White Sea and Baltic Tar Company, Mr D. W. Kelly of Carbotech and Dr S. C. Brassell for valuable discussions; Mrs A. P. Gower and Mr C. L. Saunders for assistance with GC/MS facilities, NERC for grants (GR3/2951 and GR3/3758) providing GC/MS and MS computing facilities, and British Petroleum p.l.c. for financial support (R.P.E.).

Received 9 November 1984; accepted 18 February 1985.

- Goffier, Z. *Archaeological Chemistry* (Wiley, Chichester, 1980).
- Thornton, M. D., Morgan, E. D. & Celoria, F. *Science and Archaeology* 2–3, 20–24 (1970).
- Morgan, E. D., Titus, L., Small, R. J. & Edwards, C. *Arctic* 36, 356–360 (1983).
- Rotlander, R. C. A. & Schlichtherle, H. *Archaeophysika* 10, 260–267 (1979).
- Cassar, M., Robins, G. V., Fletton, R. A. & Alstin, A. *Nature* 303, 238–239 (1983).
- Plahter, U. *Saertrykk Fra Universitetets Oldsamlings Arkiv*, 217–219 (1969).
- Mills, J. S. & White, R. *Stud. Conserv.* 22, 12–31 (1977); *Natn. Gallery tech. Bull.* 6, 3–18 (1982).
- White, R. & Castle, S. A. *Britannia* 9, 383–392 (1978).
- Simoneit, B. R. T. thesis, Univ. Bristol (1975).
- Thomas, B. R. in *Organic Geochemistry: Methods and Results* (eds Eglinton, G. & Murphy, M. T. J.) 599–618 (Springer, Berlin, 1969).
- Thomas, B. R. in *Phytochemical Phylogeny* (ed. Harborne, J. B.) Ch. 4, (Academic, New York, 1970).
- Forbes, R. J. *Bitumen and Petroleum in Antiquity*, 1–31 (Brill, Leiden, 1936).
- Danischewsky, M. I. *Post-Medieval Archaeology* Vol. 9, 77–78 (Society for Post-Medieval Archaeology, 1975).
- Meiggs, R. *Trees and Timber in the Ancient Mediterranean World* (Clarendon, Oxford, 1982).
- Jerman, K. thesis, Univ. Bristol (1984).

Filamentous microfossils from the 3,500-Myr-old Onverwacht Group, Barberton Mountain Land, South Africa

Maud M. Walsh & Donald R. Lowe

Department of Geology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

The Swaziland Supergroup, Barberton Mountain Land, South Africa, has long been regarded as a promising location for the Earth's oldest fossils because it includes some of the most ancient well-preserved sedimentary rocks, many of which contain carbonaceous matter. Although there have been numerous reports of microfossils from Swaziland Group rocks^{1–7}, the biogenicity of most of the structures has been questioned^{8–10}. Although some of the organic spheroids are probably biogenic^{10,11}, the best early Archaean simple spheroids are generally regarded as 'possible microfossils'¹⁰ because organic spheroids may form abiotically in several ways¹². The discovery of less-simple biological morphologies is therefore important in establishing the existence of early life forms in the early Archaean. Uniformly-sized curving filaments, especially tubular ones, are difficult to explain as anything other than the fossil remains of filamentous organisms. Here we report the discovery of numerous filaments from two different stratigraphical positions in the 3,500-Myr-old Onverwacht Group of the Swaziland Supergroup. Their morphologies and abundance provide convincing evidence for the existence of bacteria- or cyanobacteria-like organisms on the Earth during the early Archaean. This supports recent reports of similar filamentous microfossils from 3,500-Myr-old rocks from Western Australia¹³.

The Onverwacht Group consists of ultramafic and mafic lavas interbedded with minor felsic tuffs and sedimentary units, including carbonaceous cherts. Nd-Sm dating of komatiitic flows in both lower and upper parts of the Onverwacht Group yields an age of 3,500 Myr¹⁴. A minimum age of 3,300 Myr for

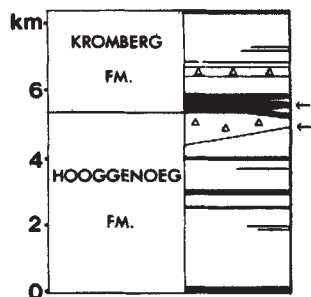


Fig. 1 Stratigraphical column of the Hooggenoeg and Kromberg Formations, upper Onverwacht Group, Swaziland Supergroup, Barberton Mountain Land. Unornamented mafic and ultramafic volcanic rocks; Δ , volcanoclastic rocks; solid shading, chert. Arrows indicate position of fossiliferous samples.

the Onverwacht Group is provided by K-Ar dating of low-grade regional metamorphism which has affected the entire sequence¹⁵.

The oldest unit in which we have identified filamentous microstructures is located near the top of the Hooggenoeg Formation in the upper half of the Onverwacht Group (Fig. 1). The structures occur in a black and white banded chert that is interbedded with thin silicified volcanoclastic layers at the base of a thick sequence of felsic tuffs and conglomerates. It crops out ~2 km north of the Komati Gorge section described by Viljoen and Viljoen¹⁶. The Hooggenoeg sample locality is 26°00'30"S, 31°00'05"E. It is exposed in a south-west-facing fault escarpment of the major ridge that generally trends north-east. A native trail crosses the ridge in a saddle south of the outcrop. Carbonaceous matter in the black chert bands occurs as fine laminae, as granules draped by laminae and as layers of irregularly shaped grains. The carbonaceous matter is identified by its petrographic characteristics—its brown-black colour and lack of crystallinity. Furthermore, whole rock analyses yield a total organic carbon content of 0.04%. Within this layer, the filamentous structures are thread-like or cylindrical and non-septate (Fig. 3a), with cross-sectional diameters ranging from <0.2 μm to 2.6 μm (Fig. 2a). The filaments are the same colour as surrounding organic matter, so they are probably partially carbonaceous. In addition, fine pyrite grains along the filaments are visible using reflected light microscopy. In a single carbonaceous layer ~2.5 cm thick, 185 individual filaments were noted in a total area of thin section 20-cm square.

The second sample containing carbonaceous filaments was collected in the immediately overlying Kromberg Formation. The filaments occur in a section of black and white banded cherts interlayered with flat- and cross-laminated clastic carbonates ~380 m above the base of the formation in the Komati Gorge section of Viljoen and Viljoen¹⁶. (The Kromberg sample locality is 26°01'42"S, 30°00'05"E. The prominent outcrop of banded chert is exposed on the path along the south bank of Komati River, ~2 km upriver from the footbridge.) The fossiliferous sample consists of alternating bands of brown-black carbonaceous chert and colourless translucent chert. The total organic carbon content of the rock is 0.11%. The carbonaceous chert layers, varying from 0.3 to 4 cm thick, have undulating tops and bottoms and are made up of billowy brown-black organic matter interlayered with fine carbonaceous laminae, both permineralized with silica. Within the fine laminae are mudstone clasts and pellets of organic matter. Many of the filaments are associated with the fine laminae, but some are within the cloudy organic matter. There are at least two distinct populations of filaments present. The most abundant filaments are threadlike (Fig. 3b, f), ranging in diameter from 0.1 to 0.6 μm and lengths from 10 to 150 μm (Fig. 2b). The other population consists of tubular filaments (Fig. 3b, c and d) with cross-sectional diameters ranging from 1.4 to 2.2 μm and lengths from 10 to 150 μm (Fig. 2b). Most filaments in this size range appear

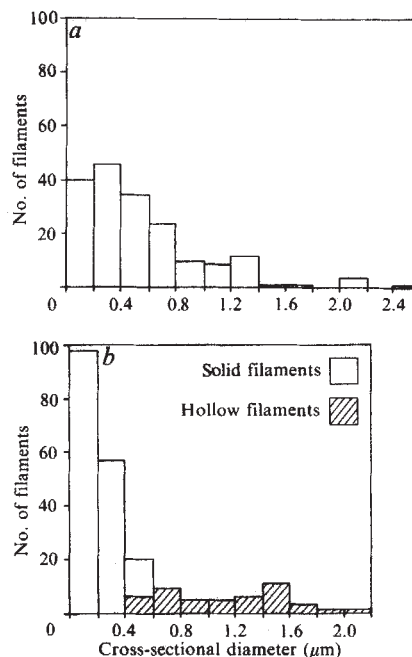


Fig. 2 Size-frequency distribution of filamentous microstructures. *a*, Hooggenoeg sample (185 microstructures); *b*, Kromberg sample (216 microstructures).

non-septate, but a few exhibit slight constrictions at intervals of ~1 μm . Based on the same transmitted and reflected microscopic criteria described for the Hooggenoeg sample, the filaments appear to be composed of carbonaceous matter, and in some cases admixed pyrite grains. The filaments are much more abundant in this sample than in the Hooggenoeg sample, with more than 200 individuals located in a 2.4-cm square area of thin section.

The age of the filaments, only slightly more than 1,000 Myr younger than the Earth itself, requires that the structures be appraised carefully with respect to their biogenicity. First, the characteristics of the rocks containing the filaments must be considered. All of the relevant features of the rocks are consistent with biogenicity of the filaments. The rock units in which the fossiliferous samples occur are clearly sedimentary units. The Hooggenoeg sample is in a 3.5-m-thick section of interbedded black and white cherts and flat- and cross-laminated sand and volcanic accretionary lapilli. The cherts are part of the sequence, with conformable contacts with over- and under-lying sand units. The sedimentary sequence overlies a komatiitic lava flow that has a brecciated weathered top. The Kromberg sample is from cherts that have a sedimentary lower contact with a pillowed lava flow. The flow has a brecciated top and signs of alteration due to submarine weathering. The section of cherts is conformably overlain by flat- and cross-laminated carbonates¹⁷. Traced laterally, the chert unit maintains its position in the sequence and is an original part of the sedimentary sequence. The rocks have undergone only low-grade metamorphism¹⁶ and thus have not been heated sufficiently to destroy carbonaceous fossils. The black or brown-black colour of the organic matter in the rocks indicates that it has experienced heating conditions consistent with low-grade metamorphism¹⁸. The similar thermal history of the carbonaceous matter and the surrounding rocks strengthens the argument for a similar time of formation for the carbonaceous rocks and the surrounding sediments and volcanic flows. Carbon isotope analyses on black-and-white-banded cherts from the same Kromberg section yield $^{13}\text{C}_{\text{PDB}}$ values of -34.5 and $-33.9 \pm 0.2\%$. These values are close to values reported for both Archaean and Proterozoic fossiliferous stromatolitic samples¹⁸. Perhaps most importantly, the microscopic texture of the carbonaceous cherts is like that of stromatolitic cherts, although larger-scale stromatolitic relief is lacking. The fine

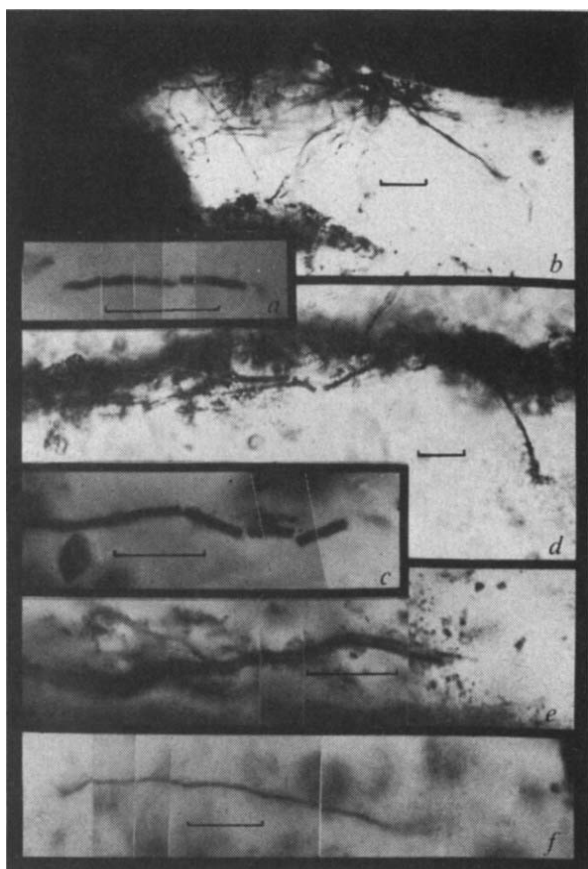


Fig. 3 Filamentous microfossils: *a*, cylindrical microfossil from Hooggenoeg sample; *b*, threadlike and tubular filaments extending between laminae, Kromberg sample; *c*, *d*, *e*, tubular filaments oriented subparallel to bedding, Kromberg sample; *f*, threadlike filament flattened parallel to bedding, Kromberg sample.

crinkly apparently-coherent carbonaceous laminae are similar to those formed by microbial mats populated by filamentous organisms that precipitate and/or trap sediment between successive growth layers.

Characteristics of the filaments themselves indicate that they are both indigenous to the original deposit and biological in origin. The filaments are located far from any weathering surfaces or veins and have a similar colour to surrounding organic matter, indicating a similar thermal history. The filaments are, therefore, unlikely to be recent endolithic organisms¹⁹. The filaments cut across crystal domains and are thus not merely concentrations of organic matter at crystal boundaries. Both the size and the shape of the carbonaceous filaments are similar to

those of modern filamentous bacteria and cyanobacteria and Proterozoic filamentous fossils interpreted to be the remains of cyanobacteria or bacteria²⁰⁻²⁴. The filaments are abundant and have a narrow size range. Each filament has a fairly constant diameter along its length. The spatial relationship of the filaments to lamination, generally either subparallel or perpendicular (Fig. 3), is like that of filaments in modern fossil mats constructed by phototrophic organisms²⁵. Many of the threadlike filaments that are aligned parallel to bedding appear ribbon-like, as if they had been compressed during sediment compaction (Fig. 3f). Such flattening of microfossils has been described from both shales²⁶ and cherts²⁷. Solid minerals or recent contaminants would be unlikely to show such flattening. The apparent composition of the filaments, carbonaceous matter and fine pyrite, is consistent with biogenicity. Filamentous microfossils consisting of pyrite grains have been reported from both the Beck Springs Dolomite²⁸ and the Gunflint Formation²¹ and haematite-coated cylindrical microfossils are found in the Belt Supergroup²⁹. Furthermore, cross-sectional views of larger filaments in the Kromberg sample show that they are hollow tubes, not solid mineral cylinders (R. J. Horodyski, personal communication. Horodyski has noted solid pyritic cylinders of abiogenic origin in cherts of the Belt Supergroup, Northwestern Montana.) The presence of pyrite along the filaments means either that the pyrite replaced the organic matter making up the walls of degrading filaments, or that the pyrite or a mineral it replaced was precipitated on the living organisms. Note that the apparent diameters of the filaments may reflect pyrite crystal size, or thickness of the pyrite coating, and not original filament size.

The filaments reported here represent the most probable biological microstructures yet reported from the Swaziland Supergroup and are similar in age to carbonaceous filamentous microfossils described from rocks in the eastern Pilbara block, Western Australia¹³. They are also the same age as the oldest known stromatolites, reported from Western Australia^{30,31}. Carbonaceous cherts with organic mat-like textures are not uncommon in the Swaziland Supergroup. Their abundance, the above described microfossils and the occurrence of both microfossils and stromatolites in Western Australia provide compelling evidence for the existence of mat-forming organisms in early Archaean shallow water environments.

This research was supported by NSF grant EAR 78-19908 to D.R.L. and an NSF graduate fellowship to M.M.W. We thank R. J. Horodyski and A. H. Knoll for guidance and encouragement. Discussions with many participants in the 1982 NASA Planetary Biology and Microbial Ecology research conference were very useful. S. M. Awramik read an early version of the manuscript. G. R. Byerly provided the Hooggenoeg sample, as well as useful criticism. Total organic carbon analyses cited were performed at the University of Cape Town. Carbon isotope analyses were provided by J. M. Hayes of Indiana University. The samples described are housed in the Louisiana State University collection of Precambrian rocks and are available for examination through D.R.L.

Received 17 September 1984; accepted 3 January 1985.

1. Pflug, H. *Econ. Geol. Res. Unit. Witwatersrand Univ. Inf. Circ. No. 28* (1966).
2. Barghoorn, E. S. & Schopf, J. W. *Science* **152**, 758-763 (1966).
3. Schopf, J. W. & Barghoorn, E. S. *Science* **156**, 508-512 (1967).
4. Nagy, B. & Nagy, L. A. *Nature* **223**, 1226-1228 (1969).
5. Engel, E. A. J. *et al. Science* **161**, 1005-1008 (1968).
6. Muir, M. D. & Grant, P. R. in *The Early History of the Earth* (ed. Windley, B. F.) 595-604 (Wiley, New York, 1976).
7. Muir, M. D. & Hall, D. O. *Nature* **252**, 376-378 (1974).
8. Schopf, J. W. *A. Rev. Earth planet. Sci.* **3**, 213-249 (1975).
9. Schopf, J. W. *Origins Life* **7**, 19-36 (1976).
10. Schopf, J. W. & Walter, M. R. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 214-239 (Princeton University Press, New Jersey, 1983).
11. Knoll, A. H. & Barghoorn, E. S. *Science* **198**, 396-398 (1977).
12. Folsome, C. A., Allen, R. D. & Ichinose, M. K. *Precamb. Res.* **2**, 263-275 (1975).
13. Awramik, S. M., Schopf, J. W. & Walter, M. R. *Precamb. Res.* **20**, 357-374 (1983).
14. Hamilton, P. J., Evensen, M. M., O'Nions, R. K., Smith, H. S. & Erlank, A. J. *Nature* **297**, 298-300 (1979).

15. DeWit, M. J., Hart, R., Martin, A. & Abbott, P. *Econ. Geol.* **77**, 1783-1802 (1982).
16. Viljoen, R. P. & Viljoen, M. J. *Geol. Soc. S. Afr. Spec. Pap.* **2**, 113-151 (1969).
17. Lowe, D. R. & Knauth, L. P. *J. Geol.* **85**, 699-723 (1977).
18. Hayes, J. M., Kaplan, I. R. & Wedeking, K. W. in *Earth's Earliest Biosphere* (ed. Schopf, J. W.) 93-134 (Princeton University Press, New Jersey, 1983).
19. Cloud, P. & Morrison, K. *Precamb. Res.* **9**, 81-91 (1979).
20. Brock, T. D. *Biology of microorganisms* (Prentice-Hall, Englewood Cliffs, 1979).
21. Barghoorn, E. S. & Tyler, S. A. *Science* **147**, 563-577 (1965).
22. Schopf, J. W. *J. Paleontol.* **42**, 651-688 (1968).
23. Awramik, S. M. & Barghoorn, E. S. *Precamb. Res.* **5**, 121-142 (1977).
24. Knoll, A. H. *J. Paleontol.* **56**, 755-790 (1982).
25. Monty, C. L. V. *Ann. Soc. Geol. Belg.* **90**, 55-100 (1967).
26. Horodyski, R. J. & Bloeser, B. *Science* **199**, 682-684 (1978).
27. Hofmann, H. J. *J. Paleontol.* **50**, 1040-1073 (1976).
28. Gutstadt, A. M. & Schopf, J. W. *Nature* **223**, 165-167 (1969).
29. Horodyski, R. J. *Precamb. Res.* **2**, 215-254 (1975).
30. Lowe, D. R. *Nature* **284**, 441-443 (1980).
31. Walter, M. R., Buick, R. & Dunlop, M. S. R. *Nature* **284**, 443-445 (1980).

Opposite motivational effects of endogenous opioids in brain and periphery

Antoine Bechara & Derek van der Kooy

Department of Anatomy, University of Toronto,
Toronto, Ontario, Canada M5S 1A8

Many psychoactive drugs, including the opiates, have been shown to have paradoxical reinforcing effects^{1,2}. Opiates produce positive reinforcing effects when they are paired with visual and textural environmental stimuli in rats^{3,4}, yet, at similar doses and over the same routes of administration, produce aversive effects, as shown when they are paired with taste stimuli⁵. Similarly, in human, the positive reinforcing effects of opiates are well known to addicts and recreational drug users, yet patients receiving opiates as analgesics often report nauseous reactions. At present there is no evidence to differentiate between the neural substrates that mediate these opposite motivational effects. We now report an initial step in the resolution of this paradox by demonstrating that endogenous and exogenous opioids produce positive reinforcing effects through an action on brain opiate receptors, and aversive effects through an action on peripheral opiate receptors (especially in the gut).

The microinjection of morphine directly into certain specific brain sites can produce positive reinforcing effects similar to those seen after systemically administered drug¹. The central dose required to produce these effects is lower than that needed to produce similar effects after systemic administration, suggesting that the primary population of receptors mediating the positive reinforcing effects is localized in the brain⁶. Recently, a large population of opiate receptors has been found on primary sensory neurones in the peripheral nervous system⁷. Local application of opiates in the periphery produces receptor-mediated hyperalgesic or irritation effects⁸. These results suggest that opiates, and by analogy endogenous opioids, produce positive reinforcing effects in the brain and aversive effects in the periphery. To test this hypothesis, we studied the motivational effects of specific opioid receptor antagonists which do or do not cross the blood-brain barrier.

Using a place-preference paradigm³, separate groups of drug-naïve rats were administered (subcutaneously (s.c.) or intraperitoneally (i.p.)) various doses of naltrexone or its quaternary derivative methyl naltrexone, which does not cross the blood-brain barrier effectively⁹. Briefly, each rat received four drug injections spaced equally over 8 days, and was immediately confined after each injection to a visually and texturally distinctive environment for 30 min. On the 4 alternating days when no opiate antagonist injection was given, each rat was injected with

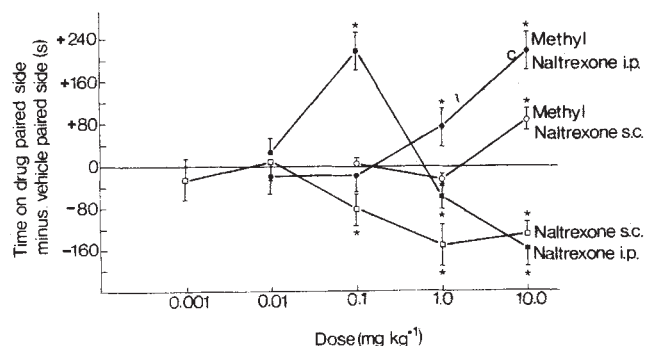


Fig. 1 Effects of various doses of naltrexone and methyl naltrexone (s.c. and i.p.) in the place-conditioning paradigm. Each point represents the mean $\pm$ s.e.m. for eight rats ($*P < 0.05$). Overall analyses of variance on the data from i.p. (0.01–10 mg kg⁻¹) and s.c. (0.1–100 mg kg⁻¹) injections revealed most importantly significant three-way interactions of time in the drug-versus-saline paired environment with drug and with dose ($F_{3,42} = 19.57$ (i.p.), $F_{2,28} = 12.91$ (s.c.), $P < 0.005$).

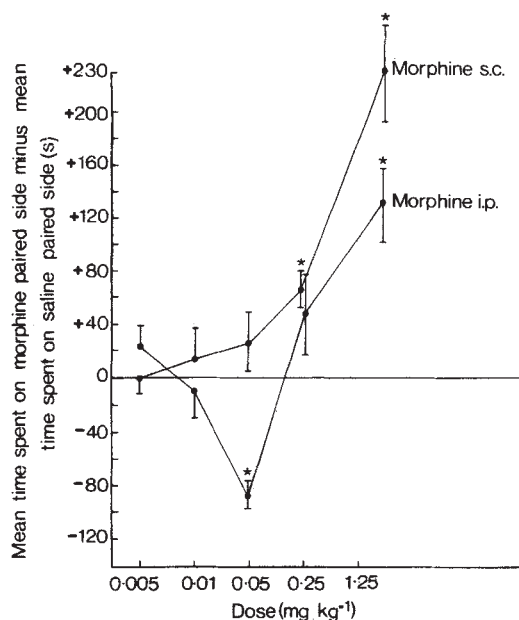


Fig. 2 Effects of various doses of morphine (s.c. and i.p.) in the place-conditioning paradigm. Each point represents the mean $\pm$ s.e.m. for eight rats ($*P < 0.05$).

saline vehicle and confined for 30 min to another separate and different distinctive environment³. Order of injections and drug-paired environment were counterbalanced within groups. On day 9, each uninjected rat was tested by recording the amount of time the rat spent in each of the two previously paired distinctive environments when given a 10-min period to explore both environments freely.

Figure 1 shows that, regardless of the route of administration, increasing doses of naltrexone were aversive whereas increasing doses of methyl naltrexone were positively reinforcing. The results suggest that the place aversions observed were due to an antagonism by the antagonist of endogenous opioid peptides acting on central opiate receptors. This suggestion is based on the fact that most opiate receptors are located in the brain and therefore the net effects of naltrexone, which can cross the blood-brain barrier, are mainly caused by its binding to central receptors. Similar results have been reported previously using another opiate antagonist, naloxone³. The preferences produced by methyl naltrexone, which does not effectively cross the blood-brain barrier, must result primarily from its binding to peripheral opiate receptors and its blockade of peripheral endogenous opioid actions.

Interestingly, one anomalous result occurred at 0.1 mg kg⁻¹ naltrexone i.p., but not s.c., which produced a place preference ($F_{1,7} = 17.23$, $P < 0.05$). We hypothesized that this was the result of a local block of aversive opiate effects in the gut without significant central action at this low dose. This implies that the peripheral aversive effects of opiates are mediated primarily by receptors in the gut. Based on this hypothesis, we predicted that a low dose of morphine, given i.p., should produce aversion. Figure 2 shows that 0.05 mg kg⁻¹ of morphine (i.p.) produced significant place aversions ($F_{1,7} = 27.47$, $P < 0.05$); the same dose administered s.c. had no effect. Higher doses of morphine (s.c. and i.p.) were positively reinforcing in the place-preference paradigm; doses below 0.05 mg kg⁻¹ had no effects. Because the vagus nerve carries sensory information from the gut to the brain, we predicted further that subdiaphragmatic vagotomies should abolish the observed effects at the low doses of i.p. morphine and naltrexone (0.05 mg kg⁻¹ morphine place aversions and 0.1 mg kg⁻¹ naltrexone place preferences), but should not abolish the opposite motivational effects seen at higher i.p. doses (1.25 mg kg⁻¹ morphine place preferences and 10.0 mg kg⁻¹ naltrexone place aversions). Figure 3 confirms these predictions. Vagotomy (verified at death by measuring

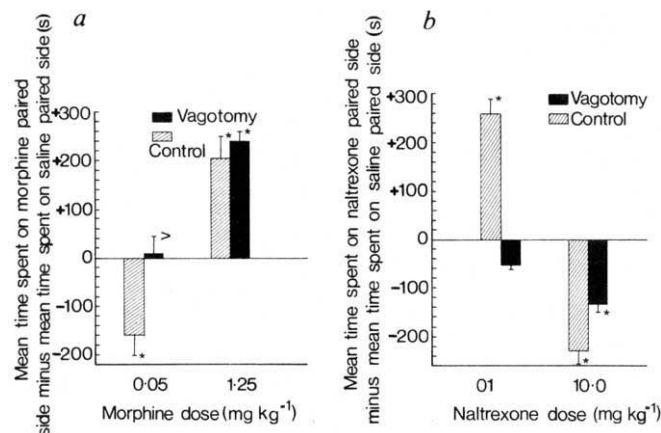


Fig. 3 *a*, Effects of vagotomy on the low-dose i.p. morphine place aversions and higher-dose i.p. place preferences. *b*, Effects of vagotomy on the low-dose i.p. naltrexone place preferences and higher-dose i.p. place aversions. Bars are means $\pm$ s.e.m. for 7–12 rats. * Indicates significant place conditioning ($P < 0.05$). Overall analyses of variance on the data from control and vagotomized animals injected with morphine and naltrexone revealed significant two-way interactions of control versus vagotomized animals with time in drug-versus-saline paired environment ($F_{1,14} = 17.88$, $P < 0.05$ (naltrexone); $F_{1,22} = 6.05$, $P < 0.05$ (morphine)).

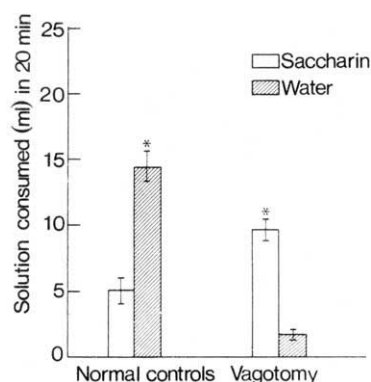


Fig. 4 Effects of vagotomy on the aversions elicited by high-dose i.p. morphine injections in the conditioned taste aversion paradigm. Bars are means $\pm$ s.e.m. for 7–10 rats. * Indicates a significant difference between saccharin and water consumption ($P < 0.05$).

food retention in the stomach¹⁰) blocked the 0.05 mg kg⁻¹ morphine-induced place aversions and the 0.1 mg kg⁻¹ naltrexone-induced place preferences. No significant effects of vagotomies were observed on the 1.25 mg kg⁻¹ morphine place preferences nor on the 10.0 mg kg⁻¹ naltrexone place aversions. These data support the hypothesis that the aversive properties of opiates are mediated by means of gut opiate receptors.

We tested this hypothesis further by investigating the effects of vagotomy on the aversive effects of opiates in the conditioned taste aversion paradigm⁵, where only aversive effects are seen even with higher doses of morphine. Fluid-deprived vagotomized ($n = 7$) and control ($n = 10$) rats received five 20-min presentations (spaced equally over 10 days) of 0.1% saccharin solution followed immediately by a morphine injection (15 mg kg⁻¹, i.p.), and on the 5 alternating days normal water paired with saline vehicle injection. Overall analyses of variance on a final drug-free, 20-min, two-bottle choice test revealed a significant two-way interaction of vagotomy versus control with saccharin versus water consumption ($F_{1,18} = 18.35$, $P < 0.05$). The control rats demonstrated an aversion for the saccharin flavour previously paired with morphine, whereas the vagotomized and morphine-injected rats showed the normal preference for saccharin (Fig. 4) seen in uninjected animals. These results again suggest that the vagus nerve carries information concerning the aversive properties of opioids from the gut to the brain.

In conclusion, endogenous and exogenous opioids acting on peripheral receptors (especially in the gut) produce aversive or nauseous effects, whereas opioids acting on central brain receptors produce positive reinforcing or euphoric effects. The present results suggest that the adverse effects of opiates experienced in human patients receiving opiates for analgesia might be minimized by blocking peripheral opiate receptors. These data also provide the first steps in analysing how the organization of the nervous system places constraints on the learning of associations between the two different, anatomically specific motivational effects of opioids and stimuli in different sensory systems.

This work was supported by the MRC of Canada. We thank Dr H. Merz for the gift of methylnaltrexone and Dr H. Weingarten for advice about vagotomies.

Received 22 November 1984; accepted 29 January 1985.

1. Cappell, H. & Leblanc, A. E. *Addictive Behav.* **1**, 55–64 (1975).
2. Wise, R. A., Yokel, R. A. & Dewitt, H. *Science* **191**, 1273–1275 (1976).
3. Mucha, R. F., van der Kooy, D., O'Shaughnessy, H. & Bucenicks, P. *Brain Res.* **243**, 91–105 (1982).
4. Weeks, J. R. *Science* **138**, 143–144 (1962).
5. van der Kooy, D. & Phillips, A. G. *Pharmac. Biochem. Behav.* **6**, 637–641 (1977).
6. van der Kooy, D., Mucha, R. F., O'Shaughnessy, M. & Bucenicks, P. *Brain Res.* **243**, 107–117 (1982).
7. Ninkovic, M., Hunt, S. P. & Gleave, J. R. W. *Brain Res.* **241**, 197–206 (1982).
8. van der Kooy, D. *Neurosci. Lett. Suppl.* **10**, S498 (1982).
9. Valentino, R. J., Herling, S., Woods, J. H., Medzhradsky, F. & Merz, H. J. *Pharmac.* **217**, 652–659 (1981).
10. Martin, J. R., Rogers, R. C., Novin, D. & van der Weele, D. A. *Bull. psychon. Soc.* **10**, 291–294 (1977).

Role of guanine nucleotide binding protein in the activation of polyphosphoinositide phosphodiesterase

Shamshad Cockcroft & Bastien D. Gomperts

Department of Experimental Pathology, University College London, London WC1E 6JJ, UK

Interaction of ligands with 'Ca²⁺-mobilizing' receptors is thought to result in the generation of two second messengers, inositol trisphosphate and diacylglycerol, from a common substrate, phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P₂) (refs 1, 2), a component of plasma membranes^{3,4}. It is not known how the occupation of such receptors is translated into the activation of the catalytic unit polyphosphoinositide (PPI) phosphodiesterase, and then to cellular activation, but our recent experiments suggest that GTP regulatory proteins may be involved. In mast cells, non-hydrolysable analogues of GTP introduced and then trapped in the cytosol are able to substitute for external ligands in inducing exocytosis, a well-defined Ca²⁺-dependent process⁵, suggesting that guanine nucleotide regulatory proteins may act by stimulating the catalytic activity of the PPI phosphodiesterase. We now provide evidence that mast cell secretion is inhibited by internalized neomycin, a compound known to interact with PPI⁶. We also show that the PPI phosphodiesterase of human neutrophil plasma membranes can be activated simply by adding GTP analogues in the presence of concentrations of Ca²⁺ that pertain in unstimulated cells. These findings strongly support the idea that the coupling factor linking receptor and PPI phosphodiesterase is a guanine nucleotide binding protein analogous to those involved in the activation and inhibition of adenylate cyclase⁷.

Mast cells, when transiently permeabilized in the presence of the GTP analogue GTP- γ -S, secrete histamine when Ca²⁺ is added subsequently to the external medium⁵. To test the possibility that PPI hydrolysis is required for the action of GTP- γ -S, we investigated the effects of neomycin, which prevents the hydrolysis of PPIs. Figure 1 shows that when neomycin at 10–100 μ M is loaded into the cytosol with GTP- γ -S (50 μ M), Ca²⁺-dependent secretion is abolished. Neomycin inhibits

Table 1 GTP- γ -S-dependent hydrolysis of polyphosphoinositides and corresponding increase in inositol phosphates

Polyphosphoinositides	Radioactivity (c.p.m.) in:	
	PtdIns(4)P	PtdIns(4,5)P ₂
Control	21,579 $\pm$ 453	19,854 $\pm$ 397
GTP- γ -S	17,362 $\pm$ 359	13,112 $\pm$ 470
Loss of label	4,217	6,742
Inositol phosphates	InsP ₂	
	InsP ₂	InsP ₃
Control	7,245 $\pm$ 155	14,450 $\pm$ 1,845
GTP- γ -S	11,746 $\pm$ 344	21,072 $\pm$ 370
Increase in label	4,501	6,672

Abbreviations: PtdIns(4)P, phosphatidylinositol 4-phosphate; PtdIns(4,5)P₂, phosphatidylinositol 4,5-bisphosphate; InsP₂, inositol bisphosphate; InsP₃, inositol trisphosphate. Human peripheral neutrophils were prepared as described previously¹⁴. The cells (10⁶ per ml) were preincubated in a balanced salt solution¹⁴ containing ³²P-phosphate (100 μ Ci ml⁻¹) for 90 min to prelabel the phospholipids, including the polyphosphoinositides. A light membrane fraction enriched in plasma membranes was prepared as described previously²⁵. The plasma membrane suspension was suspended in 20 mM PIPES buffer containing 150 mM KCl, pH 6.8, at a protein concentration of 3–4 mg ml⁻¹. Samples (500 μ l) were transferred to tubes containing GTP- γ -S and Ca-EGTA buffer (3 mM EGTA final) in an equal volume to give final concentrations of GTP- γ -S and Ca²⁺ of 100 μ M and 100 nM, respectively. After 10 min at 37 °C, the reaction was terminated by the addition of 3.75 vol chloroform/methanol/concentrated HCl (100:200:5). This mixture was allowed to stand for 30 min and then separated into two phases by the addition of 1.25 vol each of chloroform and water. The lower (chloroform) phase was retained for lipid analysis and the upper (aqueous) phase retained for the determination of inositol phosphates. Phospholipids were separated by TLC on oxalate-impregnated silica gel-H-coated plates using chloroform/methanol/acetone/acetic acid/water (40:13:15:12:8) and the radioactive spots located by autoradiography. Areas containing PtdIns4P and PtdIns(4,5)P₂ were scraped and the lipids hydrolysed in 72% perchloric acid at 180 °C for 1 h. The sample was made up to 10 ml in water and counted by Cerenkov radiation. For the analysis of the water-soluble inositol phosphates, the upper (aqueous) phase was neutralized, made up to 10 ml with water and run on Dowex 1 (formate form) as described previously²⁶. Results (c.p.m.) are expressed as the means ($\pm$ s.d.) of triplicate determinations from a single experiment. Similar results were obtained in two other experiments. Samples of labelled membranes were also treated with Ca²⁺ (1 mM) alongside each experiment for comparison and on this particular occasion, Ca²⁺-dependent activation of the phosphodiesterase caused loss of PtdIns4P (55%) and PtdIns(4,5)P₂ (65%) compared with GTP- γ -S-activated loss of PtdIns4P (20%) and PtdIns(4,5)P₂ (34%).

secretion only when presented to the cells together with ATP⁴⁻, which causes cell permeabilization⁸ and so permits the drug entry to the cytosol. This result suggested that PPI hydrolysis is an integral step in the activation of cells loaded with non-hydrolysable GTP analogues and that PPI phosphodiesterase is subject to control by a guanine nucleotide binding protein. To test this possibility, we used the PPI phosphodiesterase of human neutrophils. We have described recently a plasma membrane PPI phosphodiesterase in human and rabbit neutrophil membranes which can be activated in the presence of Ca²⁺ (ref. 9). Like the red cell enzyme¹⁰, the neutrophil enzyme is not under the control of cytosol Ca²⁺ because Ca²⁺ is required in the millimolar range. The possibility must therefore be considered that the enzyme is controlled by some other so far unidentified component.

GTP- γ -S can activate the neutrophil enzyme (Table 1) in the conditions of Ca²⁺ homeostasis pertaining in resting cells. The hydrolysis of PPI yields the water-soluble products inositol bis- and trisphosphate. The loss of label from PPI is fully accounted for by the appearance of label in the inositol phosphates (Table 1). Data for Ca²⁺ activation of the enzyme are also included for comparison (see Table 1 legend). Our earlier experiments included Mg²⁺ (1 mM) in the incubation mixtures, but it soon became clear that this leads to hydrolysis of the inositol phosphates, with the implication that Mg²⁺-requiring phosphatases

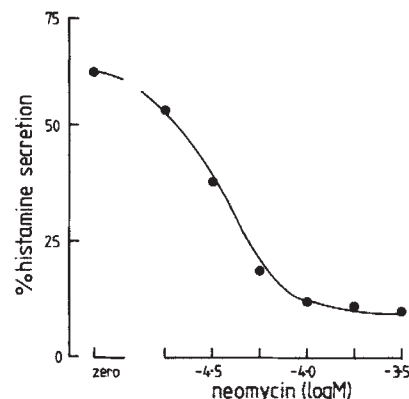


Fig. 1 Inhibition by neomycin of Ca²⁺-dependent histamine secretion from GTP- γ -S-loaded mast cells. Exogenous solutes were loaded into the cytosol of rat mast cells as described previously⁵; they were treated with ATP (3.16 μ M), GTP- γ -S (50 μ M) and neomycin (concentrations as indicated on the abscissa). This treatment with ATP in the absence of divalent cations causes the formation of lesions in the plasma membrane and allows the other solutes present to penetrate into the cytosol. At 6 min, MgCl₂ (5 mM) was added to close the membrane lesions and so trap the loaded GTP- γ -S and neomycin in the cytosol. After 15 min further incubation, exocytotic secretion was triggered by the addition of Ca²⁺ (5 mM). The cells were quenched with ice-cold phosphate-buffered saline 10 min later, and secreted histamine was measured as described previously⁸. Neomycin at concentrations up to 300 μ M had no effect on the histamine assay. The results illustrated are representative of four independent experiments.

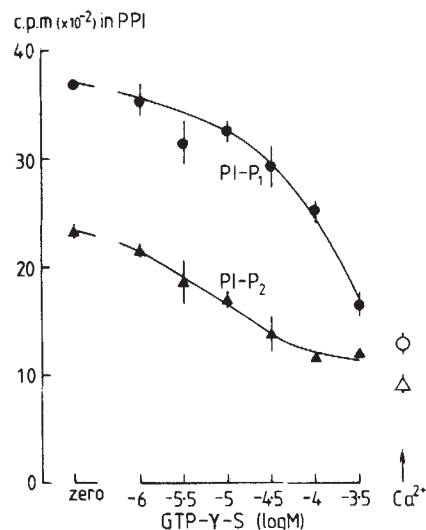


Fig. 2 Concentration-effect relationship of GTP- γ -S-activated polyphosphoinositide phosphodiesterase in neutrophil membranes. Experimental details are the same as outlined in Table 1 legend except that the incubations were performed using 100 μ l of membrane suspension added to 100 μ l of buffer containing Ca-EGTA (100 nM Ca²⁺), 1 mM Mg²⁺ (final) and GTP- γ -S as indicated. For comparison, data for PPI phosphodiesterase activation by 1 mM Ca²⁺ are also shown (open symbols). The results illustrated are representative of four separate experiments. Results are expressed as c.p.m. in the individual polyphosphoinositides; determinations were made in triplicate. Error bars denote s.d. Circles denote PtdIns4P (PI-P₁) and triangles denote PtdIns(4,5)P₂ (PI-P₂).

are present; such phosphatases have been identified recently in the red cell¹¹. Mg²⁺ was therefore excluded subsequently from experiments in which water-soluble derivatives were to be analysed. There was no indication that Mg²⁺ is required for the action of GTP- γ -S on the enzyme, and it has little effect on the inositol lipid phosphomonoesterase⁹ and no effect on PPI phosphodiesterase⁹.

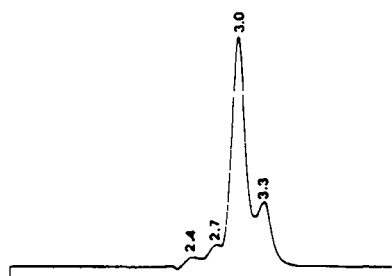


Fig. 3 HPLC analysis of commercial GTP- γ -S. GTP- γ -S (10 μ M of 1 mM) was injected into a Bondapak cartridge-type column and the sample run isocratically as described previously^{3,14}. Chart recording speed was set at 20 mm min⁻¹. GTP- γ -S and IDP were obtained from Boehringer Mannheim and all other nucleotides from Sigma. GppNHp was obtained from both sources and produced similar HPLC profiles. Elution times (min) for the various nucleotides were: ATP, 4.95; GTP, 2.98; GDP, 3.27; GMP, 4.47; ADP, 5.7; GTP- γ -S, 3.0 (68%) and 3.3 (20%, probably GDP); GppNHp, 3.07 (65%) and 3.42 (30%); GppCH₂p, 3.04; IDP, 3.32.

Activation of PPI phosphodiesterase was dependent on the concentration of GTP- γ -S in the range 1–300 μ M (the highest concentration tested) (Fig. 2). Activity was already apparent at 1 μ M and at higher concentrations approached activities that can be achieved by millimolar concentrations of Ca²⁺ (ref. 9). However, we stress that these results refer to the added compound, obtained from commercial sources. The guanine nucleotides are inherently unstable; in particular, there is a tendency towards deamination to yield inosine homologues¹². Analysis by HPLC reveals that GTP- γ -S (Fig. 3) and GppNHp (see Fig. 3 legend) are always very impure; the impurity detected is most probably GDP in the case of GTP- γ -S (Fig. 3) and remains unidentified in the case of GppNHp. The proportions of the individual components did not alter during 8 months' storage. However, the HPLC method^{13,14} does not resolve guanine from their equivalent inosine nucleotides, and so we cannot be certain of the actual concentrations required for activation of the enzyme, although we do know that the onset of activation occurs at concentrations <1 μ M. In five separate experiments, 1 μ M GTP- γ -S hydrolysed 10% ($\pm$ 2) (s.e.m., n = 5) of phosphatidylinositol 4-phosphate and PtdIns(4, 5)P₂.

We also tested the specificity of the GTP analogues, but the uncertainties involved in the actual concentrations present prevented us from quantifying their relative potencies. All compounds were tested at a nominal 100 μ M. GDP, GDP- β -S, ATP, ATP- γ -S and AppNHp were inactive, whereas GTP- γ -S, GppNHp and GppCH₂p caused activation. The effect of GTP was erratic, probably because of hydrolysis by a nonspecific nucleotidase known to be present in neutrophil plasma membranes¹⁵.

The results described here provide firm evidence that PPI phosphodiesterase can be activated by analogues of GTP, and so by inference may be subject to control by the binding of GTP at a receptor-controlled guanine nucleotide binding protein (N-protein) in its cellular environment (Fig. 4). In this it is similar to adenylate cyclase, although there are also important differences. In particular, whereas the availability of substrate to adenylate cyclase is effectively inexhaustible, the substrate for PPI phosphodiesterase is an endogenous membrane lipid, strictly limited by the inositol lipid kinases, and thus perhaps under independent control (S.C., J. Lamb and E. Zanders, in preparation). Further, there are two products of phosphodiesterase action, one of which (diacylglycerol) is retained in the organic phase of the membrane, while the other (inositol trisphosphate) is released into the cytosol.

Binding of ligand to receptor leads to the binding of GTP and activation of N-protein. Different receptors are likely to interact with separate but related families of N-proteins; N₁, N₂ and transducin have already been described¹⁶. We designate the

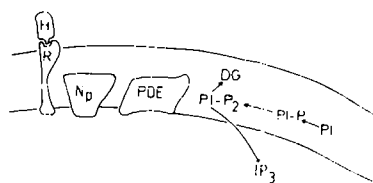


Fig. 4 Schematic description of events that occur when hormone binds to its receptor and results in transmembrane signalling. Interaction of hormone (H) with receptor (R) allows a guanine nucleotide binding protein (N_p) to exchange bound GDP for GTP. This generates an activated species of N_p which activates the catalytic unit, polyphosphoinositide phosphodiesterase (PDE). The hydrolysis of phosphatidylinositol bisphosphate (PI-P₂) generates two second messengers—diacylglycerol (DG) to activate protein kinase C and inositol trisphosphate to release Ca²⁺—which then together mediate cellular responses. The supply of the substrate (PI-P₂) is limited by the activity of the inositol lipid kinases which sequentially phosphorylate phosphatidylinositol (PI) to phosphatidylinositol 4-phosphate (PI-P) and then to PI-P₂.

putative N-protein which is involved in the activation of PPI phosphodiesterase as N_p; an inhibitory counterpart to N_p analogous to N_i probably also exists. The activated N_p is then responsible for switching on the PPI phosphodiesterase as shown in Fig. 4.

These ideas are further supported by the recent finding that islet activating protein (purified from pertussis toxin) can inhibit mast cell and neutrophil secretion induced by both internalized GTP analogues¹⁷ and receptor-directed agonists¹⁸. This suggests that N_p is susceptible to ADP-ribosylation and thus inhibition similar to N_i and transducin.

The precedent for involving N-proteins was set initially by the adenylate cyclase system⁷ and this role has now been extended not only to hormone receptor action, but also to the biosynthesis of proteins¹⁹. Other examples include the light-induced activation of cyclic GMP phosphodiesterase²⁰ and the activation of protein kinase and cyclic AMP phosphodiesterase by insulin^{21,22}. The protein products of the *ras* oncogene family also have a binding site for GTP, again implying a role for N-protein^{23,24}. Thus, the activation of PPI phosphodiesterase shares with other receptor and non-receptor signal-transducing systems a common mechanism of control involving a family of related but specific N-proteins.

This work was supported by grants from the Vandervell Foundation, the Wellcome Trust and the MRC.

Received 18 December 1984; accepted 5 February 1985.

1. Michell, R. H. *Trends pharmac. Sci.* **8**, 263–265 (1983).
2. Berridge, M. J. *Biochem. J.* **220**, 345–360 (1984).
3. Downes, P. & Michell, R. H. *Cell Calcium* **3**, 467–502 (1982).
4. Cockcroft, S., Taylor, J. & Judah, J. D. *Biochim. biophys. Acta* (in the press).
5. Gomperts, B. D. *Nature* **306**, 64–66 (1983).
6. Schacht, J. J. *Lipid Res.* **19**, 1063–1067 (1978).
7. Rodbell, M. *Nature* **284**, 17–22 (1980).
8. Bennett, J. P., Cockcroft, S. & Gomperts, B. D. *J. Physiol., Lond.* **317**, 335–345 (1981).
9. Cockcroft, S., Baldwin, J. M. & Allan, D. *Biochem. J.* **221**, 477–482 (1984).
10. Downes, C. P. & Michell, R. H. *Biochem. J.* **202**, 53–58 (1982).
11. Downes, C. P., Mussat, M. C. & Michell, R. H. *Biochem. J.* **203**, 169–177 (1982).
12. *Remarks Concerning the Stability of GDP, GTP and dGTP* (Boehringer Mannheim technical leaflet, 1984).
13. Anderson, Q. S. & Murphy, R. C. *J. Chromat.* **121**, 251–262 (1976).
14. Cockcroft, S. *Biochim. biophys. Acta* **795**, 37–46 (1984).
15. Hyslop, P. A. *et al.* *FEBS Lett.* **166**, 165–169 (1984).
16. Houslay, M. D. *Trends biochem. Sci.* **9**, 39–40 (1984).
17. Nakamura, T. & Ui, M. *FEBS Lett.* **173**, 414–418 (1984).
18. Okajima, F. & Ui, M. *J. biol. Chem.* **259**, 13863–13871 (1984).
19. Kaziro, Y. *Biochim. biophys. Acta* **505**, 95–127 (1978).
20. Stryer, L. *Cold Spring Harb. Symp. quant. Biol.* **48**, 841–852 (1983).
21. Heyworth, C. M., Rawal, S. & Houslay, M. D. *FEBS Lett.* **154**, 87–91 (1983).
22. Heyworth, C. M., Wallace, A. V. & Houslay, M. D. *Biochem. J.* **214**, 99–110 (1983).
23. McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. *Nature* **310**, 644–649 (1984).
24. Newbold, R. *Nature* **310**, 628–629 (1984).
25. Bennett, J. P., Caswell, A. H., Cockcroft, S. & Gomperts, B. D. *Biochem. J.* **208**, 801–808 (1982).
26. Downes, C. P. & Michell, R. H. *Biochem. J.* **198**, 133–140 (1981).

Antigen-specific interaction between T and B cells

Antonio Lanzavecchia

Basel Institute for Immunology, Grenzacherstrasse 487,
Postfach CH-4005, Basel, Switzerland

It is well known that B cells require T-cell help to produce specific antibody. Classic experiments suggested that antigen-specific helper T cells interact with antigen-specific B cells via an antigen 'bridge'^{1,2}, the B cells binding to one determinant on an antigen molecule (the 'hapten'), while the T cells at the same time recognize another determinant (the 'carrier'). T-helper cells bind specifically to antigen-presenting cells (APC), which have picked up and processed the appropriate antigen, and this interaction, like the interaction of T-helper cells with specific B cells, is restricted by products encoded by the major histocompatibility complex (MHC)³⁻⁵. Whereas conventional APC such as macrophages display no binding specificity for antigen, B cells have clonally distributed antigen-specific surface immunoglobulin receptors which would be expected to enhance their capacity to present antigen to T cells^{6,7}. These findings are difficult to reconcile with the simple 'antigen bridge' mechanism of interaction, because it is hard to visualize how the bimolecular complex (processed antigen plus MHC molecule) on the APC surface can resemble the trimolecular complex (antigen bound to surface immunoglobulin plus MHC molecule) on the B-cell surface. To address this problem, we have cloned and immortalized human antigen-specific B cells with Epstein-Barr virus (EBV) and analysed their interaction with T-cell clones specific for the same antigen. We report here that surface immunoglobulin is indeed involved in the uptake and concentration of antigen, allowing specific B cells to present antigen to T cells with very high efficiency. However, the antigen must first be internalized and processed by specific B cells and it is then presented to T cells in an MHC-restricted manner indistinguishable from that characteristic of conventional APC.

Three B-cell clones producing IgG antibody against tetanus toxoid (TT) were isolated by EBV-induced transformation of peripheral blood B cells from an immune donor. The capacity of these B cells to bind antigen was tested by saturation analysis using ¹²⁵I-labelled TT. The double reciprocal plot of the bound versus free antigen showed an affinity constant of 5×10^7 – 2×10^8 l mol⁻¹, the number of binding sites varying from 2×10^4 to 10^5 per cell (see Table 1). The affinity for antigen of the secreted antibody was comparable within a factor of 3 to that of the B cells (data not shown). By contrast, B cells cloned from the same donor but producing IgG of an irrelevant specificity bound only very limited amounts of TT and showed no tendency for saturation. Furthermore, binding of ¹²⁵I-TT to TT-specific B cells was inhibited by unlabelled TT whereas binding to non-specific B cells was not (Table 1). These data indicate that specific B cells bind soluble antigen with a high affinity equal to the affinity of the antibody they produce, while B cells not specific for TT (nonspecific) bind much lower, although still measurable, amounts of antigen in an immunoglobulin-nonspecific way.

We compared the ability of TT-specific and nonspecific B cells to present TT or an unrelated antigen to T cells. The dose of antigen in culture was titrated for any T/B-cell combination and the T-cell proliferative response was assessed by thymidine incorporation. We found that both specific and nonspecific B cells were able to present native TT to TT-specific T-cell clones (Fig. 1a–c). However, while presentation by nonspecific B cells required high concentrations of antigen in culture (Fig. 1c), TT-specific B cells could trigger T-cell proliferation in the presence of 10^4 -fold lower antigen concentrations (10^{-11} – 10^{-12} M, Fig. 1a,b). We have obtained similar results in 12 T/B-cell combinations using 4 different T-cell clones and 3 different specific B-cell clones.

Table 1 Antigen binding by antigen-specific B cells

B-cell clone	¹²⁵ I-TT bound per 10 ⁶ cells (c.p.m.)	k (l mol ⁻¹)	No. of sites per cell
11.3	43,529 (1,039)	5×10^7	10^5
8.5	26,834 (593)	2×10^8	2×10^4
4.2	21,067 (912)	5×10^7	5×10^4
7.1	841 (730)	ND	ND

TT was labelled with ¹²⁵I by the chloramine-T method. B cells from three TT-specific EBV-transformed B-cell clones (11.3, 8.5 and 4.2) or from a clone producing an immunoglobulin of irrelevant specificity (7.1) were incubated with increasing concentrations of ¹²⁵I-TT in 200 µl phosphate-buffered saline containing 0.5% bovine serum albumin and 15 mM Na₂S₂O₃. After 1 h on ice, the cells were pelleted and 100 µl of supernatant was collected. The cells were resuspended, washed twice by centrifugation through a cushion of fetal calf serum (FCS) and the radioactivity bound was determined. c.p.m. represent the amount of ¹²⁵I-TT bound near the saturation point of the curve. Numbers in parentheses indicate c.p.m. bound at the same concentration of labelled TT in the presence of a 100-fold excess of cold TT. k, The affinity of TT binding by B cells, and the number of binding sites were calculated from the double reciprocal plot of the saturation curve as described previously¹². ND, the value was not determined because saturation could not be achieved.

Table 2 Antigen presentation by specific B cells requires intracellular antigen processing

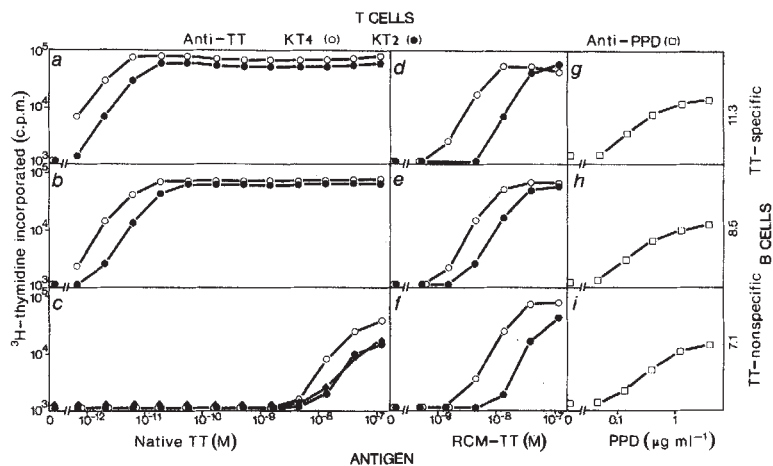
		³ H-thymidine incorporated (c.p.m.) by T-cell clone		
Expt 1: Treatment of B cells before washing	Added in culture	KT2	KT4	AK3
—	—	161	138	32,625
TT	—	8,578	36,853	33,499
Anti-Ig, TT	—	979	6,985	33,029
Anti-Ig, TT	TT	8,552	36,491	ND
TT, anti-Ig	—	5,984	28,616	ND
TT	Anti-Ig	8,425	33,890	ND
Expt 2: Treatment of B cells before fixation				
—	—	127	214	21,779
—	TT	148	171	22,532
TT	—	12,373	79,175	21,708
Chloroquine + TT	—	220	962	21,582

Expt 1: A TT-specific B-cell clone (8.5) was washed, resuspended in RPMI/FCS and incubated on ice with native TT (5×10^{-10} M) or with a rabbit anti-human F(ab')₂ affinity-purified antibody ($300 \mu\text{g ml}^{-1}$). After 15 min, anti-F(ab')₂ or anti-TT were added as indicated for 1 h. The cells were then washed three times, irradiated (3,000 R) and cultured with TT-specific T cells as described in Fig. 1 legend. Native TT (5×10^{-10} M) or anti-immunoglobulin ($300 \mu\text{g ml}^{-1}$) were added where indicated. Expt 2: A TT-specific B-cell clone (11.3) was washed, resuspended in RPMI 10% FCS and pulsed at 37 °C with 10^{-9} M native TT or medium alone in the presence or absence of 10^{-5} M chloroquine. After 4 h the cells were washed four times with Hank's balanced salt solution (HBSS) and fixed with 0.05% glutaraldehyde for 1 min at room temperature. The reaction was stopped with 0.2 M glycine in HBSS and the cells were washed three times. Glutaraldehyde-fixed and fresh B cells bound, respectively, 44,331 and 31,520 c.p.m. of ¹²⁵I-TT per 10⁶ cells. Fixed cells (5×10^4) were cultured in 96-well U-bottom microplates with 2×10^4 T cells from TT-specific clones (KT2 and KT4) or an alloreactive clone (AK3) specific for K.K. (the TT-immune donor) as a control. Soluble TT was added in culture at 10^{-7} M where indicated. Thymidine incorporation was measured as in Fig. 1. Data represent the geometric mean of triplicate determinations.

As a specificity control, we compared the ability of TT-specific and nonspecific B cells to present two antigens not recognized by their specific surface antibodies. The first was the denatured form of TT itself. As generally found for complex protein antigens⁸, denaturation of TT followed by reduction and carboxymethylation results in an antigen preparation that has lost the

Fig. 1 Antigen presentation by antigen-specific B lymphocytes. Cells (2×10^4) from the TT-specific T-cell clones KT2 (●) and KT4 (○) or from a PPD-specific T-cell line (□) were cultured with 2×10^4 irradiated (3,000 R), EBV-transformed B-cell clones specific for TT (a, d, g, clone 11.3; b, e, h, clone 8.5) or non-TT-specific (c, f, i, clone 7.1) in 200 μ l RPMI/FCS in 96-well flat-bottom microplates. The cultures were stimulated with a wide range of concentrations of native TT, reduced carboxymethylated-TT (RCM-TT) or PPD. In c, clone KT2 was also stimulated in the presence (◆) of $3 \mu\text{g ml}^{-1}$ purified antibody produced by clone 11.3. After 2 days the cultures were pulsed for 16 h with $1 \mu\text{Ci}$ ^3H -thymidine and the incorporated radioactivity was counted. The background of irradiated B cells alone (always $<3,000$ c.p.m.) was subtracted.

Methods. Antigens: Native TT was obtained from the Massachusetts Public Health Biological Laboratories and purified by gel filtration on Sephadex G-200. RCM-TT was prepared by denaturation of native TT in 6 M urea, followed by reduction with dithiothreitol and carboxymethylation with iodoacetic acid. The anti-TT antibody produced by specific B cells (see below) was purified from the culture supernatant and coupled to CNBr-activated Sepharose. This immunoadsorbent was used to deplete RCM-TT of contaminating antigen still reactive with these antibodies. This antigen preparation contained $<0.1\%$ antigen still reactive with antibody as detected by inhibition of binding of ^{125}I -TT. PPD was obtained from Serum Institut, Copenhagen. T cells: T-cell clones specific for TT were derived as described previously¹¹. Briefly, peripheral blood mononuclear cells (PBM) from a TT-immune donor (K.K.) were stimulated *in vitro* with 10^{-7} M TT. After 7 days, the activated T cells were expanded in medium supplemented with 30 U ml^{-1} recombinant interleukin-2 (IL-2, Biogen) and 2 weeks later restimulated with TT and autologous irradiated (3,000 R) PBM. After a further expansion in IL-2, the cells were cloned by limiting dilution in the presence of irradiated autologous PBM, TT and IL-2. Five TT-specific clones were isolated and maintained in culture for more than 8 months by periodic restimulations with antigen, followed by expansion in IL-2. All the T cells used in the present experiments were obtained from 3–4-week-old IL-2-dependent cultures and were virtually devoid of contaminating APC, as they were unable to proliferate in response to soluble antigen unless appropriate APC were added. PPD-specific T cells from donor K.K. were also obtained from PBM stimulated with PPD. The cells were selected by three cycles of restimulation and grown in IL-2 as described above. B cells: TT-specific B-cell clones were obtained from donor K.K. by EBV transformation of peripheral blood B cells essentially as described elsewhere¹³. PBM (10^7) were resuspended in 10 ml RPMI/FCS containing 30% supernatant of the EBV-producing marmoset cell line B95.8 and 600 ng ml^{-1} cyclosporin A (Sandoz). After 16 h at 37°C , the cells were washed and seeded at 5×10^4 cells per well in 96-well flat-bottom microplates in 200 μ l RPMI/FCS containing 600 ng ml^{-1} cyclosporin A. After 13 days the supernatants were tested for the presence of IgG anti-TT antibodies by enzyme-linked immunosorbent assay¹¹. Cells from eight positive wells were cloned and subcloned by limiting dilution in the presence of 3×10^5 rat thymocytes as filler cells. Three different TT-specific clones producing IgG anti-TT antibody (0.5 – $5 \mu\text{g ml}^{-1}$) were isolated as well as several other clones producing IgG not specific for TT.



ability to react with antibodies while maintaining the capacity to stimulate T cells. Figure 1d–f shows that there is no significant difference in the low efficiency with which TT-specific and nonspecific B cells present the denatured form of TT. As a further control, an unrelated antigen, purified protein derivative of tuberculin (PPD), was used. No difference was found between TT-specific and nonspecific B cells in their capacity to present PPD to a PPD-specific T-cell line (Fig. 1g–i).

To test the possibility that secreted antibodies contribute to the high efficiency of antigen presentation by specific B cells, monoclonal anti-TT IgG was purified from the culture supernatants of the B-cell clones and added at different concentrations to cultures containing TT-specific T-cell clones and nonspecific B cells. In no case was any significant increase in the efficiency of presentation of TT observed (Fig. 1c). From the above experiments, we conclude that surface immunoglobulin molecules have a role in antigen uptake and concentration on specific B cells, allowing them to cooperate preferentially with specific T cells in the presence of minute amounts of antigen.

To investigate whether specific T/B-cell cooperation, like antigen presentation by macrophages or nonspecific B cells^{3,5}, is MHC-restricted, TT-specific B cells were compared for their ability to present TT to autologous or allogeneic TT-specific T-cell clones. TT-specific B cells were unable to present TT to allogeneic T cells even in the presence of concentrations of antigen 10^4 times higher than those required for presentation to autologous T cells (Fig. 2a). Furthermore, monoclonal anti-HLA-DR antibody quantitatively inhibited antigen presentation (Fig. 2b). We conclude, therefore, that these antigen-specific T/B-cell interactions are MHC-restricted.

Macrophages and nonspecific B cells must usually first internalize and process a complex protein antigen before presenting it to T cells^{4,6}. However, it was possible that specific B cells could bypass this step and present protein antigens bound externally to immunoglobulin receptors. We therefore tested the effect of anti-immunoglobulin antibodies on antigen uptake and

presentation by specific B cells. The addition of anti-immunoglobulin to B cells before pulsing with limiting concentrations of TT effectively reduced their subsequent capacity to stimulate T cells (Table 2, expt 1), the same treatment resulting in an 80% inhibition of uptake of ^{125}I -TT (data not shown). However, when anti-immunoglobulin was added after TT, no inhibition was observed, indicating that these antibodies interfere only with the uptake but not with the subsequent antigen presentation to T cells. We next tested the effect of chloroquine treatment, which inhibits intracellular antigen processing⁹, followed by glutaraldehyde fixation to prevent any further processing¹⁰. Table 2 (expt 2) shows that glutaraldehyde-fixed TT-specific B cells were able to present native TT to TT-specific T cells only if the B cells had been pulsed with TT before fixation. Unpulsed glutaraldehyde-fixed B cells did not present soluble TT, even though they could still bind ^{125}I -TT. If chloroquine was added during the antigen pulse, the ability of TT-pulsed B cells to stimulate T cells was completely abrogated. None of the above treatments, however, affected the capacity of B cells to present an alloantigen to a specific alloreactive T-cell clone. These data indicate that a chloroquine-sensitive intracellular step is required for antigen presentation and that the role of surface immunoglobulin is limited to antigen uptake.

Our finding that specific B cells can present TT to T cells at free antigen concentration $\sim 10^4$ times lower than required for presentation by nonspecific B cells or peripheral blood monocytes¹¹ agree with results reported by Rock *et al.* using a hapten-carrier conjugate and hapten-specific normal B cells⁷. These experiments provide direct quantitative evidence for the role of surface immunoglobulin in concentrating antigen on specific B cells and suggest that this may be the basis of clonal selection in the specific antibody response to low antigen concentrations. Furthermore, our data obtained with a monovalent protein antigen show that selective interaction can occur at antigen concentrations far lower than the affinity constant of the antibody, that is, in conditions where only a very small

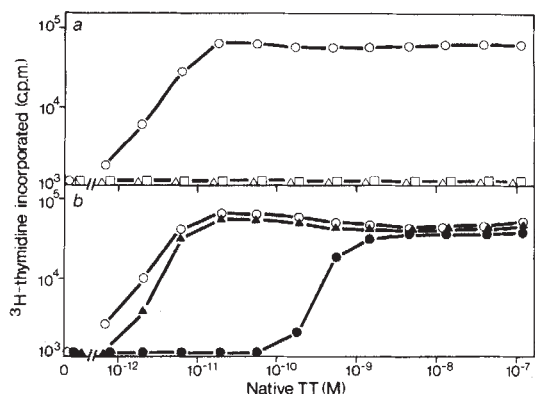


Fig. 2 Antigen presentation by antigen-specific B cells is MHC-restricted. *a*, A TT-specific B-cell clone (11.3) was irradiated and cultured in the presence of different concentrations of TT with an autologous TT-specific T-cell clone, KT1 (○), or with allogeneic TT-specific clones AT1 (□) and FT4 (Δ). The latter clones, obtained as described in Fig. 1 legend from two donors unrelated to K.K., incorporated in parallel experiments >30,000 c.p.m. when stimulated by TT in the presence of autologous APC (data not shown). *b*, TT-specific T and B cells (clones KT4 and 11.3) were stimulated with TT in the presence of anti-HLA-DR monoclonal antibody D1.12 at $10 \mu\text{g ml}^{-1}$ (●), anti-HLA-A,B,C antibody W6/32 at $50 \mu\text{g ml}^{-1}$ (▲) or medium alone (○). Culture conditions were as described in Fig. 1 legend.

fraction of surface immunoglobulin is occupied by antigen. It is possible that receptor internalization, by continuously removing antigen from the equilibrium, could lead with time to selective antigen accumulation in specific B cells.

The above experiments assayed T/B-cell interaction by measuring T-cell activation. We also have evidence that B-cell responses, such as increased antibody production and induction of terminal B-cell differentiation, can occur in the same system at the same antigen dose range. Thus, despite the limitations inherent in the use of EBV-transformed B cells and long-term T-cell clones, we feel that these data on antigen presentation are representative of T/B-cell cooperative interaction.

The requirement for antigen processing by specific B cells allows, in principle, all T cells selected by macrophage-processed antigen to interact with the same specificity with all antigen-specific B lymphocytes, independently of the hapten epitope recognized by the B cell. The requirement for processing may vary with the form of antigen used for stimulation¹⁰, however, since the B-cell response is focused mainly against conformational determinants⁸, which are easily lost on processing, it is likely that antigen processing by B cells will be a quite general rule. Antigen processing by B cells also easily explains the linked cooperation between two cells of different specificities. B cells bind soluble antigens by recognizing easily accessible conformational epitopes and, after intracellular processing of the antigen, will present to the T cells a different series of determinants similar to those produced by any nonspecific APC. This suggests that the antigen bridge is due to a sequential, rather than simultaneous, recognition of antigen by B and T cells.

I thank Sylvie Bove for technical assistance, Walter Fiers and Biogen for providing human recombinant IL-2, Lynn Gisler for preparation of the manuscript, Polly Matzinger and Jonathan Howard for criticism and suggestions, Fritz Melchers and Franco Celada for discussion and support. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co., Basel, Switzerland.

Received 12 November 1984; accepted 21 January 1985.

- Mitchison, N. A. *Eur. J. Immunol.* **1**, 18–27 (1971).
- Rajewsky, K., Schirmacher, W., Nase, S. & Jerne, N. K. *J. exp. Med.* **129**, 1131–1136 (1969).
- Rosenthal, A. S. & Shevach, E. M. *J. exp. Med.* **138**, 1194–1212 (1973).
- Unanue, E. R., Beller, D. I., Lu, C.Y. & Allen, P. M. *J. Immunol.* **132**, 1–5 (1984).
- Katz, D. H., Hamaoka, M. E. & Benacerraf, B. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2624–2628 (1973).
- Kakiuki, T., Chestnut, R. W. & Grey, H. M. *J. Immunol.* **131**, 109–114 (1982).

- Rock, K. L., Benacerraf, B. & Abbas, A. K. *J. exp. Med.* **160**, 1102–1113 (1984).
- Benacerraf, B. *J. Immunol.* **120**, 1809–1812 (1978).
- Ziegler, H. K. & Unanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 175–178 (1982).
- Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. *J. exp. Med.* **158**, 303–316 (1983).
- Lanzavecchia, A., Parodi, B. & Celada, F. *Eur. J. Immunol.* **13**, 733–738 (1983).
- Celada, F., Macario, A. J. L. & Conway De Macario, E. *Immunonchemistry* **10**, 797–804 (1973).
- Rosen, A., Persson, K. & Klein, G. *J. Immunol.* **130**, 2899–2902 (1983).

Absence of growth by most receptor-expressing fetal thymocytes in the presence of interleukin-2

Harald von Boehmer, Andrea Crisanti, Pawel Kisielow & Werner Haas

Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland

The growth of mature T cells is regulated by receptors for interleukin-2 (IL-2) and by IL-2 itself. Binding of antigen to T-cell antigen receptors induces the expression of IL-2 receptors, and binding of IL-2 to these receptors induces transferrin receptor expression and is sufficient to promote the growth of T cells for several days^{1,2}. However, nothing is known about the growth requirements of pre-T cells. We have therefore studied the dividing population of T-cell precursors which carry the Thy-1 surface antigen, but lack surface antigens Ly2 and L3T4; these cells are present in 14-day-old embryonic thymus^{3,4}. If the thymus is removed at this stage and placed in organ culture, all lymphocyte subpopulations normally present in thymuses of adult mice develop *in vitro*^{4,5}, that is, the nonfunctional Ly2⁺, L3T4⁺ population and the functional Ly2⁺, L3T4⁺ and Ly2⁺, L3T4⁺ populations⁴. We now report that, in contrast to their progeny, the early Ly2⁺, L3T4⁺ cells express large amounts of IL-2 receptors, but most of them do not grow in IL-2-containing media outside the thymus. In contrast to dividing mature T cells, most fetal thymocytes express low amounts of transferrin receptors.

We have analysed the expression of receptors for IL-2 and transferrin using monoclonal antibody 7D4, directed against IL-2 receptors⁶, and antibody 34D2, directed against transferrin receptors. The 34D2 hybridoma was obtained in our laboratory, and the antibodies precipitate a disulphide-linked dimer, consisting of subunits of relative molecular mass 93,000 (93K) and inhibit the binding of radiolabelled transferrin to the EL4 thymoma (A.C., unpublished). Table 1 shows the proportion of IL-2-receptor-positive thymocytes at various stages of thymus development; >90% of 14-day-old embryonic thymocytes were IL-2-receptor positive whereas only about 5% of adult thymocytes were stained. Moreover, 15-day-old embryonic thymocytes which showed very strong staining with antibody 7D4, stained only weakly with the anti-transferrin receptor antibody 34D2 in contrast to cells from the IL-2-dependent line CTL-L16 (Fig. 1). As analysed by gel electrophoresis, antibody 7D4 precipitated the same molecular species from both the fetal thy-

Table 1 The proportion of IL-2-receptor-positive thymocytes at various stages of thymus development

Age	7D4 positive
14-Day embryonic	90%
15-Day embryonic	72%
16-Day embryonic	20%
21-Day postnatal	4%

Thymocytes from various donors were incubated with 7D4 antibodies, washed and stained with a fluorescein isothiocyanate (FITC)-labelled sheep anti-mouse immunoglobulin (kindly provided by Luciana Forni) which cross-reacts on rat antibodies. Stained cells were analysed by fluorescence-activated cell sorter (FACS) as described elsewhere⁴ and shown in Fig. 1.

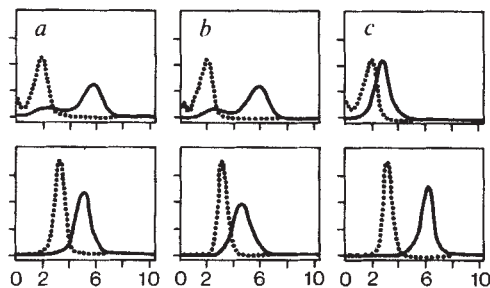


Fig. 1 Staining of 15-day-old embryonic thymocytes (top) and CTL-L16 cells (bottom) with anti-IL-2-receptor antibody 7D4 (hybridoma supernatant undiluted (a) and diluted 1:10 (b)) and anti-transferrin-receptor antibody 34D2 (undiluted supernatant (c)). Cells were incubated with antibodies at 0°C, washed and stained with FITC-sheep anti-mouse immunoglobulin. Broken line, second antibody only.

mocytes and the CTL-L16 line (Fig. 2). We did not detect any IL-2 production by 15-day-old embryonic thymocytes⁴ and, despite their expression of IL-2 receptors, fetal thymocytes did not grow outside the thymus in IL-2-containing media (Table 2).

The biological significance of IL-2 receptor expression by pre-T cells is obscure. As a first step towards an understanding, we may compare the results presented here with the data available on the induction and function of IL-2 receptors in mature T cells. First, IL-2 receptor expression by mature T cells is controlled by antigen receptors which are, at least in part, formed by two different ~45K polypeptide chains linked by disulphide bridges^{7,8}. Although 15-day-old embryonic thymocytes do not express these antigen receptor molecules, which are detectable in two-dimensional gel analysis⁹, they nevertheless express high levels of IL-2 receptors (Fig. 1). At present, however, we cannot exclude the possibility that only one of the α , β or γ polypeptide chains controls IL-2 receptor expression in fetal thymocytes. Second, IL-2, which induces growth in IL-2-receptor-expressing mature T cells, does not induce DNA synthesis in IL-2-receptor-expressing pre-T cells. Third, proliferating mature T cells express more transferrin receptors than do most fetal thymocytes. The low level of transferrin receptor expression by fetal thymocytes may be due to the lack of IL-2, which is required for transferrin receptor expression by mature T cells. The expression of IL-2 receptors in the absence of IL-2 may, in fact, be a signal for cell death and related to the finding that most newly formed thymocytes die *in situ*¹⁰.

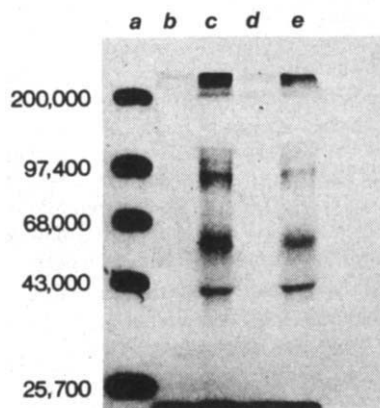


Fig. 2 Precipitation of the antibody 7D4 from lysates of CTL-L16 cells and 15-day-old embryonic thymocytes (pre-T cells). SDS-polyacrylamide gel electrophoresis (10%, reducing conditions) of material precipitated by antibody 7D4. The CTL-L16 line as well as 15-day-old embryonic thymocytes were labelled with ¹²⁵I using lactoperoxidase. NP40 lysates were precipitated with 7D4 or normal rat serum and rabbit anti-rat immunoglobulin. Lane a, relative molecular mass markers; lane b, CTL-L16 and rat serum; lane c, CTL-L16 + 7D4; lane d, pre-T cells + rat serum; lane e, pre-T cells + 7D4.

Table 2 ³H-thymidine incorporation by fetal thymocytes and CTL-L16

IL-2	15-Day-old embryonic thymocytes	CTL-L16
—	6,700	9,200
1,000 U ml ⁻¹	6,900	115,400

Cells (10⁵) were suspended in culture medium with or without IL-2 and ³H-thymidine was added immediately. Incorporation of thymidine was measured after 24 h. After 48 h no viable thymocytes were detected. Recombinant IL-2 was obtained from Biogen¹¹.

We thank Dr E. Shevach and collaborators for the gift of antibody 7D4. P.K. is on leave of absence from the Ludwik Hirsfeld Institute of Immunology and Experimental Therapy of the Polish Academy of Sciences, Wrocław, Poland. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co., Basel, Switzerland.

Received 2 October 1984; accepted 4 February 1985.

1. Larsson, E. L. & Coutinho, A. *Nature* **280**, 239–241 (1979).
2. Devos, R., Plaetnick, G. & Fiers, W. *Eur. J. Immun.* (in the press).
3. Jenkinson, E. J., Owen, J. J. T. & Aspinall, R. *Nature* **284**, 177–179 (1980).
4. Kisielow, P., Leiserson, W. & von Boehmer, H. *J. Immun.* **133**, 1117–1123 (1984).
5. Robinson, J. M. & Owen, J. J. T. *Clin. exp. Immun.* **27**, 322 (1977).
6. Malek, T. R., Robb, R. J. & Shevach, E. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5694–5698 (1983).
7. Haskins, K. *et al. J. exp. Med.* **157**, 1149–1169 (1983).
8. Acuto, O., Meuer, S. C., Schlossman, J. C. & Reinherz, E. L. *J. exp. Med.* **158**, 1368–1373 (1983).
9. Snodgrass, H. R., Kisielow, P., Kiefer, M., Steinmetz, M. & von Boehmer, H. *Nature* **313**, 592–595 (1985).
10. McPhee, D., Pye, J. & Shortman, K. *Thymus* **1**, 151–156 (1979).
11. Devos, R. *et al. Nucleic Acids Res.* **11**, 4307–4323 (1983).

Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion glycoproteins and unexpected relation to leukocyte interferon

Timothy A. Springer*, David B. Teplow† & William J. Dreyer†

* Laboratory of Membrane Immunochimistry, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

† Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

Cell-surface adherence reactions are fundamental to the biology of lymphocytes, monocytes and granulocytes. The lymphocyte function-associated 1 (LFA-1) and macrophage 1 (Mac-1) glycoproteins mediate differing types of adhesion reactions on these cells. LFA-1 participates in T-lymphocyte and natural killer cell adhesion to target cells^{1–3}, whereas the Mac-1 antigen is identical to the complement receptor type 3, which mediates adhesion of monocytes and granulocytes to C3bi-sensitized particles⁴. Deficiency of these proteins, in a heritable disease, results in multiple adhesion-related leukocyte defects^{5,6}. LFA-1 and Mac-1 resemble one another in overall structure, having α -subunits of relative molecular mass (M_r) 180,000 and 170,000, respectively, which are non-covalently associated with β -subunits of M_r 95,000 in $\alpha_1\beta_1$ complexes^{7,8}. Peptide mapping and immunological cross-reactivity have shown that the β -subunits are highly related if not identical, but have revealed no similarities between the α -subunits^{7,9,10}. Nonetheless, the shared β -subunit suggested that LFA-1 and Mac-1 might be members of a protein family containing diversified but evolutionarily related α -subunits. Therefore, we examine here the structure of the α -subunits by N-terminal amino-acid sequencing. Sequence homology shows that the α -subunits are members of a novel leukocyte adhesion protein family, and suggests that their evolution occurred by gene duplication. A search for similarities to previously sequenced proteins reveals a further unexpected homology between LFA-1 and leukocyte (α) interferons.

LFA-1	1	Asn	Leu	Asp	Thr	Arg	Pro	Thr	Gln	Ser	10	Phe	Leu	Ala	Gln	Ala	(Gly	Arg	His)	19	Phe
Mac-1	(Phe)	Asn	Leu	Asp	Thr	Glu	His	(Pro)	Met	Thr	Phe	Glx	Glu	Asn	Ala	Lys	(Gly)	Phe			

Fig. 1 N-terminal sequences of LFA-1 and Mac-1 α -subunits. Homologous sequences are boxed.

Methods. LFA-1 was purified from 10- or 50-g batches of EL-4 T lymphoma cells grown *in vitro*. Cells were lysed with Triton X-100, sodium deoxycholate was added before the ultracentrifugation step and LFA-1 antigen was bound to M17/4 monoclonal antibody-Sepharose and eluted at high pH, as described previously⁸. The eluate in 0.5% Triton X-100, 20 mM triethylamine pH 11.0, was neutralized by adding 0.075 vol. of 1 M Tris-HCl pH 6.8. Mac-1 from P388D₁ cells was purified similarly on M1/70 monoclonal antibody-Sepharose, except that the neutralized eluate was subjected to a second cycle of affinity purification. SDS-polyacrylamide gel electrophoresis (PAGE) and Coomassie brilliant blue staining of purified LFA-1 and Mac-1 demonstrated that the α - and β -subunits had no contaminants, as shown previously^{7,10}. Purified antigens were reduced with 20 mM dithiothreitol, alkylated with 50 mM iodoacetic acid and precipitated overnight at -20 °C with 1.5 vols 1% acetic acid in methanol. Precipitates were collected at 200,000 g for 0.5 h and subjected to preparative SDS-PAGE. Subunits were localized by Coomassie brilliant blue staining, electroeluted¹⁷ and subjected to gas/liquid solid phase microsequencing¹⁸. Phenylthiohydantoin amino acids were determined by HPLC¹⁹. The sequence assignments shown were confirmed by multiple determinations. The α -subunit of LFA-1 from two different purifications was sequenced three times (200, 44 and 22 μ g) and the α -subunit of Mac-1 was sequenced twice (63 and 31 μ g). Residues in parentheses were found together with smaller amounts of another residue, or in low yield. No sequence was obtained from the LFA-1 and Mac-1 β -subunits, suggesting that they are blocked.

We purified murine LFA-1 and Mac-1 antigens by monoclonal antibody affinity chromatography from detergent lysates of EL-4 T-lymphoid and P388D₁ macrophage-like cell lines and separated the α - and β -subunits (see Fig. 1 legend). The N-terminal sequences of the LFA-1 and Mac-1 α -subunits are 33% homologous (Fig. 1). Four contiguous amino acids are identical at positions 2-5 and further identities occur at positions 11 and 15. Some of the differences represent relatively conservative amino-acid substitutions, namely Phe for Tyr at residue 1, Thr for Ser at 10, and Asp for Glu at 14. If a gap was introduced before the Lys at position 16 in Mac-1, it would be aligned with an Arg in LFA-1 and Phe at positions 18 and 19 would be aligned.

The sequence homology shows that the α -subunits constitute a novel adhesion protein family, the genes of which evolved by duplication of a primordial ancestral gene. The N-terminal sequences have diverged considerably since the primordial gene duplication event. These differences suggest that the specificity in adhesive interactions is determined by the α -subunits, in agreement with localization of a binding site of the C3bi antigen on the α -subunit of Mac-1 using α - and β -subunit-specific monoclonal antibodies^{10,11}. Duplication and specialization of the α -subunits must have preceded divergence of mouse and

Table 1 Homology between LFA-1, Mac-1 and interferons

	Mac-1	% Identity			
		Rodent IFN- α	Human IFN- α	IFN- β	IFN- γ
LFA-1	33	32	35	26	5
Mac-1	—	15	7	3	6
Rodent IFN- α	—	—	55	20	0
Human IFN- α	—	—	—	27	1
Murine and human	—	—	—	—	5

The numbers refer to the percentage homology between the sequences depicted in Figs 1 and 2. Homologies between residues 7-25 of the mouse and human interferons were calculated and averaged in the groups shown.

human, as LFA-1 and Mac-1 homologues mediating the same functions occur in humans¹². A third type of human α -subunit has been identified¹² which may represent a further member of the α -subunit protein family. The amino-acid sequences described here allow the specification of oligonucleotide probes, which should be of use in the isolation and further study of the Mac-1/LFA-1 family of α -subunit genes.

A computer search¹³ for homologies with other proteins reveals an unexpected, strong homology between LFA-1 and human leukocyte interferon (Fig. 2). Multiple genes for α -interferon (IFN- α) code for a family of closely related interferon proteins¹⁴. Comparison of LFA-1 with 12 mouse, rat and human α -interferons reveals homology of LFA-1 with all of them (Fig. 2). Note that the LFA-1 and IFN- α sequences align without any gaps. IFN- α shows ~30% homology to fibroblast interferon (IFN- β). LFA-1 shows a lesser degree of homology with IFN- β (Fig. 2). The homologies with IFN- α are highly statistically significant (Fig. 2).

Homology with LFA-1 occurs in different positions in rodent and human IFN- α s. Despite the higher homology of LFA-1 to human IFN- α 1 than to any other interferon, an additional homology found exclusively with rodent IFN- α occurs at residues 1 and 2 in LFA-1 (Fig. 2). Also, conserved substitutions occur at many positions, including 10 and 11, where no identities are found. These findings further strengthen the homology between LFA-1 and IFN- α .

The homology among LFA-1, Mac-1 and the interferons is summarized in Table 1. Mac-1 and LFA-1 are about as homologous with one another as is LFA-1 with IFN- α . The significance of the 32-35% homology between LFA-1 and IFN- α s is emphasized by the fact that it is stronger than the 20-27% homology observed between IFN- α and IFN- β in the segments compared with LFA-1, or the homology of 29% along the total lengths of IFN- α and IFN- β ¹⁵. Furthermore, the homology between LFA-1 and the α - and β -interferons is stronger than the homology between IFN- γ and the α - and β -interferons.

Our findings suggest that the LFA-1/Mac-1 α -subunit family and the interferon family arose from a single ancestral gene and can be considered as a superfamily. This is remarkable for

Fig. 2 Sequence homology between LFA-1 and interferons. Residues homologous between interferons and LFA-1 are boxed; homologies with Mac-1 are italicized. α -, β - and γ -interferon sequences^{14,20,21} were aligned as described previously^{22,23}. Asp 25 of IFN- β was deleted for alignment²². The right-hand column describes the percentage probability that homology results from chance alone, as determined by summation of the log odds of each amino-acid pair derived from the mutation data matrix of Dayhoff²⁴.

		1	5	10	15	19	P (%) ^b															
LFA-1	(1-19)	Tyr	Asn	Leu	Asp	Thr	Arg	Pro	Thr	Gln	Ser	Phe	Leu	Ala	Gln	Ala	Gly	Arg	His	Phe	-	
Mouse IFN- α 1	(7-25)	His	Asn	Leu	Arg	Asn	Lys	Arg	Ala	Leu	Thr	Leu	Leu	Val	Gln	Met	Arg	Arg	Leu	Ser	1.2	
Mouse IFN- α 2	(7-25)	Tyr	Asn	Leu	Arg	Asn	Lys	Arg	Ala	Leu	Lys	Val	Leu	Ala	Gln	Met	Arg	Arg	Leu	Pro	0.3	
Rat IFN- α	(7-25)	His	Asn	Leu	Arg	Asn	Lys	Arg	Val	Phe	Thr	Leu	Leu	Ala	Gln	Met	Arg	Arg	Leu	Ser	2.0	
Human IFN- α 1	(7-25)	His	Ser	Leu	Asp	Asn	Arg	Arg	Thr	Leu	Met	Leu	Leu	Ala	Gln	Met	Ser	Arg	Ile	Ser	0.08	
Human IFN- α 2	(7-25)	His	Ser	Leu	Gly	Ser	Arg	Arg	Thr	Leu	Met	Leu	Leu	Ala	Gln	Met	Arg	Lys	Ile	Ser	0.6	
Human IFN- α 4	(7-25)	His	Ser	Leu	Gly	Asn	Arg	Arg	Ala	Leu	Ile	Leu	Leu	Ala	Gln	Met	Gly	Arg	Ile	Ser	0.08	
Human IFN- α 5	(7-25)	His	Ser	Leu	Ser	Asn	Arg	Arg	Thr	Leu	Met	Ile	Met	Ala	Gln	Met	Gly	Arg	Ile	Ser	0.16	
Human IFN- α 6	(7-25)	His	Ser	Leu	Gly	His	Arg	Arg	Thr	Met	Met	Leu	Leu	Ala	Gln	Met	Arg	Arg	Ile	Ser	0.4	
Human IFN- α 7	(7-25)	His	Ser	Leu	Arg	Asn	Arg	Arg	Ala	Leu	Ile	Leu	Leu	Ala	Gln	Met	Gly	Arg	Ile	Ser	0.13	
Human IFN- α 8	(7-25)	His	Ser	Leu	Gly	Asn	Arg	Arg	Ala	Leu	Ile	Leu	Leu	Ala	Gln	Met	Arg	Arg	Ile	Ser	0.5	
Human IFN- α 9	(7-25)	His	Ser	Leu	Asn	Asn	Arg	Arg	Thr	Leu	Met	Leu	Met	Ala	Gln	Met	Arg	Arg	Ile	Ser	0.5	
Human IFN- α c	(7-25)	His	Ser	Leu	Gly	Asn	Arg	Arg	Ala	Leu	Ile	Leu	Leu	Gly	Gln	Met	Gly	Arg	Ile	Ser	0.1	
Mouse IFN- β	(10-29)	Glu	Arg	Thr	Ser	Asn	Ile	Arg	Lys	Cys	Gln	Glu	Leu	Leu	Glu	Gln	Leu	Gly	Lys	Ile	Asn	3.2
Human IFN- β	(10-29)	Gln	Arg	Ser	Ser	Asn	Phe	Gln	Cys	Gln	Lys	Leu	Leu	Trp	Gln	Leu	Gly	Arg	Leu	Glu	126	

several reasons. First, LFA-1 and Mac-1 are surface receptors, whereas the interferons are secreted mediators. Second, there are striking differences in size. The LFA-1 and Mac-1 α -polypeptide chains (M_r 180,000 and 170,000) are roughly 10 times longer than those of the interferons ($M_r \sim 20,000$). Despite the difference in size, the N-terminal sequence of LFA-1 aligns only six residues away from the N-terminus of IFN- α . A common evolutionary mechanism for obtaining large proteins is tandem duplication and joining of domains. Further sequencing of LFA-1 is needed to determine whether it contains the homologue of an entire interferon domain and whether it contains duplications of these domains.

Received 13 November 1984; accepted 7 January 1985.

1. Davignon, D., Martz, E., Reynolds, T., Kürzinger, K. & Springer, T. A. *J. Immun.* **127**, 590-595 (1981).
2. Springer, T. A., et al. *Immun. Rev.* **68**, 111-135 (1982).
3. Pierres, M., Goridis, C. & Golstein, P. *Eur. J. Immun.* **12**, 60-69 (1982).
4. Beller, D. I., Springer, T. A. & Schreiber, R. D. *J. exp. Med.* **156**, 1000-1009 (1982).
5. Anderson, D. C. et al. *J. clin. Invest.* **74**, 536-551 (1984).
6. Springer, T. A., Thompson, W. S., Miller, L. J., Schmalstieg, F. C. & Anderson, D. C. *J. exp. Med.* **160**, 1901-1918.
7. Kürzinger, K., Ho, M. K. & Springer, T. A. *Nature* **296**, 668-670 (1982).
8. Kürzinger, K. & Springer, T. A. *J. biol. Chem.* **257**, 12412-12418 (1982).
9. Trowbridge, I. S. & Omary, M. B. *J. exp. Med.* **154**, 1517-1524 (1981).
10. Sanchez-Madrid, F., Simon, P., Thompson, S. & Springer, T. A. *J. exp. Med.* **158**, 586-602 (1983).
11. Wright, S. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5699-5703 (1983).

It is tempting to speculate that the LFA-1/Mac-1/interferon superfamily could have functional as well as structural interrelationships. All of these molecules function in cell interactions of the host defence system. Unlike the immunoglobulin superfamily¹⁶, none of them is involved in specific antigen recognition and their importance in host defence could have pre-dated the evolution of specific immunity.

We thank Dr R. Doolittle for homology searches with the NEWAT data base and Mr S. Thompson for technical assistance in protein purification. This work was supported by NIH grants CA-31798, CA-31799, and GM-06965 and an American Cancer Society faculty award to T.A.S.

12. Sanchez-Madrid, F., Nagy, J., Robbins, E., Simon, P. & Springer, T. A. *J. exp. Med.* **158**, 1785-1803 (1983).
13. Doolittle, R. F. *Science* **214**, 149-159 (1981).
14. Weissmann, C. et al. *Phil. Trans. R. Soc. B* **299**, 7-28 (1982).
15. Taniguchi, T. et al. *Nature* **285**, 547-549 (1980).
16. Williams, A. F. *Nature* **308**, 12-13 (1984).
17. Hunkapiller, M. W., Lujan, E., Ostrander, F., Hood, L. E. *Meth. Enzym.* **91**, 227-236 (1983).
18. Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. *J. biol. Chem.* **256**, 7990-7997 (1981).
19. Hunkapiller, M. W. & Hood, L. E. *Meth. Enzym.* **91**, 486-493 (1983).
20. Higashi, Y. *J. biol. Chem.* **258**, 9522-9529 (1983).
21. Gray, P. W. & Goeddel, D. V. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5842-5846 (1983).
22. Sternberg, M. J. E. & Cohen, F. E. *Int. J. biol. Macromolec.* **4**, 137-144 (1982).
23. Epstein, L. B. *Nature* **295**, 453-454 (1982).
24. Dayhoff, M. O. & Hunt, L. T. *Meth. Enzym.* **91**, 524-545 (1983).

Isolation and characterization of a cDNA clone encoding the murine homologue of the human 20K T3/T-cell receptor glycoprotein

Peter van den Elsen, Beth-Ann Shepley, Mildred Cho & Cox Terhorst

Laboratory of Molecular Immunology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

The antigen receptor on the surface of human T lymphocytes, which consists of a heterodimer of relative molecular mass (M_r) 90,000 (90K) (α - and β -chains), is associated with the T3 antigen ($\gamma = 25K$, $\delta = 20K$ and $\epsilon = 20K$)¹⁻⁷. A working model for the mode of action of the T3/T-cell receptor complex is that the clonotypic α - and β -chains are involved in the recognition and binding of antigen in the context of polymorphic major histocompatibility complex (MHC) gene products on the surface of target cells. Antigen binding by the clonotypic receptor probably results in conformational changes in this structure which are recognized by and subsequently trigger the associated T3 complex to transmit signals into the cell, resulting in a proliferative response⁸⁻¹⁰. The similarity in structure between murine and human clonotypic antigen receptors¹¹⁻¹⁵ suggests that such a mechanism of recognition and activation also exists in mouse T lymphocytes, but so far there has been no evidence for the existence of a murine T3 complex. Here we demonstrate the existence of a T3 δ -chain mRNA in murine T lymphocytes. Our sequence data strongly suggest that this mouse mRNA codes for a complete T3 δ polypeptide chain and reveal some interesting properties of the protein.

The cDNA clone pGBC-9, which contains a full-length cDNA copy coding for the human T3 δ -chain¹⁶, was used here to examine a series of murine T-cell lines for expression of RNA coding for the murine equivalent of the T3 δ -chain; only in mouse T cells was a 0.7-kilobase (kb) mRNA observed (data not shown). From one of these cells, B3C6, an anti-Ia T-T hybridoma,¹⁷ we isolated poly(A)⁺ mRNA, which was used to construct a cDNA library in the bacteriophage λ gt10 (refs 18, 19). This library was then screened with the cDNA insert of the human T3 δ cDNA clone pGBC-9 at low stringency. Ten plaques out of $\sim 2.5 \times 10^5$ were found to hybridize with pGBC-9. On plaque purification and rescreeing at low stringency,

only two plaques remained positive. The cDNA inserts of these two overlapping phages were 200 and 640 base pairs (bp) long, respectively. The 640-bp insert of phage λ PEM-T3 δ was isolated and subcloned into the *EcoRI* site of the plasmid vector SP65 (ref. 20) (that is, cDNA clone pPEM-T3 δ). The nucleotide sequence of clone pPEM-T3 δ was determined by the Maxam-Gilbert method²¹. Figure 1 shows the sequencing strategy and the resulting 570-nucleotide-long sequence of the cDNA insert of pPEM-T3 δ . Only one open reading frame is present in this sequence, which ends with a termination codon, TAA, at position 490. No initiation codon was detected. Comparison of the size of the cDNA insert with the size of the T3 δ mRNA detected in a Northern blot (data not shown) and comparison with the human T3 δ cDNA (Fig. 1b) indicates that this cDNA does not represent a full-length transcript of the murine T3 δ mRNA. Alignment with the human cDNA sequence (Fig. 1b) and the presence of a polyadenylation signal, AATAAA, followed by a poly(A) tail 19 nucleotides downstream, shows that pPEM-T3 δ contains the entire 3' noncoding region and lacks some of the 5' sequences.

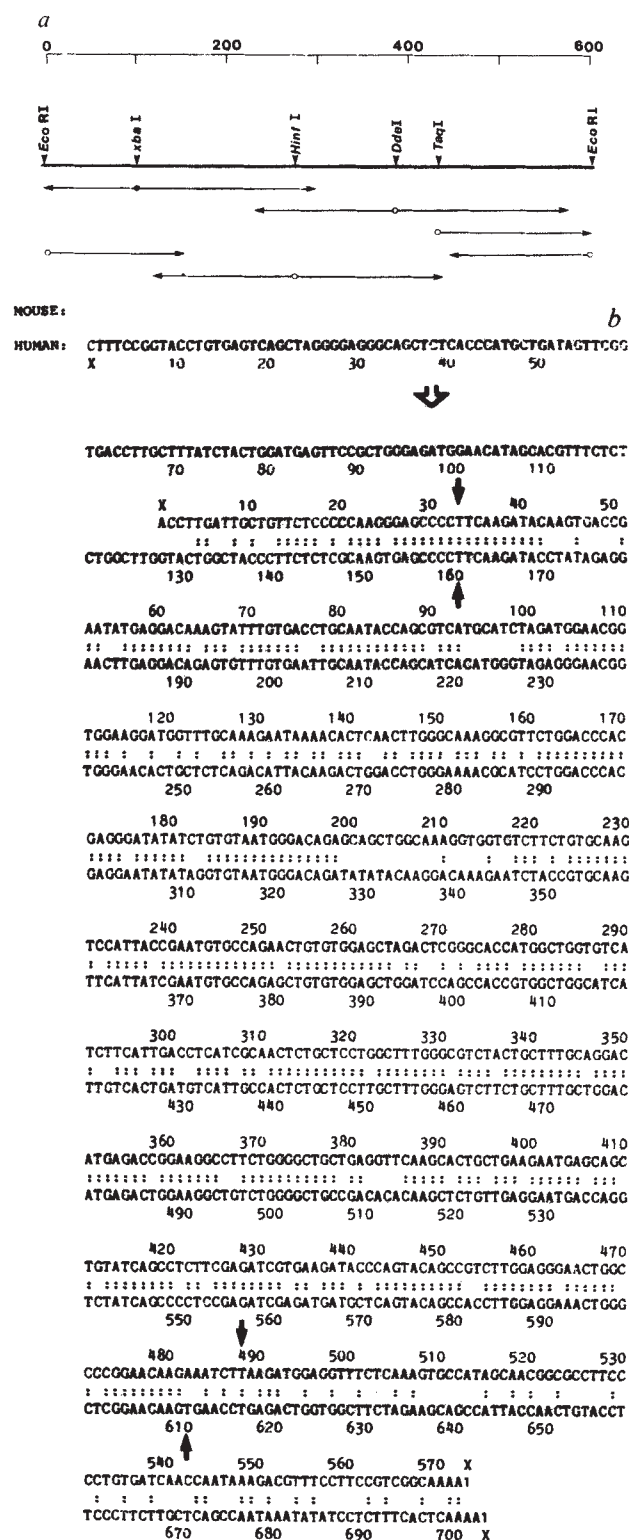
We have shown previously that in human the T3 δ mRNA is present in cells of the T lineage only¹⁶. On Northern blot analysis with the murine clone pPEM-T3 δ cDNA, mRNA was detected in T cells and thymocytes, but not in L-cells, liver cells or a B-cell hybridoma (data not shown). This suggests that the murine T3 δ -chain also is expressed exclusively in thymus-derived lymphocytes.

The PROTALN and NUCALN (IntelliGenetics, Inc.) programs²² revealed no homology between the murine T3 δ -chain and T-cell receptor structures, immunoglobulin or Thy-1. In addition, a search of the Dayhoff database using the Wilbur-Lipman algorithm^{22,23} did not reveal homology with any known protein.

Comparison of the amino-acid sequences of the human and murine T3 δ polypeptide chains showed that they are very similar. Alignment of the amino-acid sequences revealed that the murine T3 δ cDNA clone (pPEM-T3 δ) codes for only part of the leader sequence (Fig. 2). Furthermore, the mature protein has two additional amino acids at its C-terminus (Lys and Ser). The mature murine T3 δ -chain therefore consists of 152 amino acids with a deduced M_r of 16,905 and is slightly larger than its human equivalent (M_r 16,719). The domain organization of the murine T3 δ -chain, derived from its hydropathicity profile²⁴, is very similar to that of the human T3 δ -chain (Fig. 3). A 79-residue extracellular domain is followed by a transmembrane

Fig. 1 a, Partial restriction map of the murine T3 δ -chain and strategy for sequencing of clone pPEM-T3 δ . The restriction enzymes used were *Eco*RI, *Xba*I, *Hinf*I, *Dde*I and *Taq*I. Open circles represent 5'-labelling; the closed circle represents 3'-labelling for Maxam-Gilbert sequence analysis. **b**, Comparison of the nucleotide sequences of murine and human T3 δ -chain. The 570-nucleotide-long sequence of clone pPEM-T3 δ (mouse T3 δ) is compared with the 700-nucleotide-long sequence of clone pPGBC-9 (human T3 δ). For the alignment we used the NUCALN (IntelliGenetics, Inc.) program. Identical nucleotides are indicated by double dots. The ATG initiation codon in the human T3 δ sequence, and the codon coding for the first amino acid of the mature protein (Phe) and termination codon of both human and murine T3 δ are indicated by arrows.

Methods. The murine B3C6 cDNA library constructed in phage λ gt10 was screened with the cDNA insert of clone pPGBC-9. Approximately 2.5×10^5 plaques were generated and plaque lifts were made on nitrocellulose filters as described elsewhere²⁷. The filters were prehybridized at 42°C in 50% deionized formamide, $5 \times$ SSC ($1 \times$ SSC = 0.15 M NaCl, 0.015 M Na₃-citrate), $5 \times$ Denhardt's solution ($1 \times$ Denhardt = 0.02% Ficoll 400, 0.02% bovine serum albumin and 0.02% polyvinylpyrrolidone), 50 mM sodium phosphate (pH = 6.5), 0.1% SDS and 250 μ g ml⁻¹ of denatured salmon sperm DNA. Hybridization was carried out in the same buffer for 16 h at 42°C. The human T3 δ -chain cDNA insert of clone pPGBC-9 (ref. 16), ³²P-labelled by nick-translation (specific activity $2-5 \times 10^6$ c.p.m. μ g⁻¹), was used as hybridization probe at 5×10^6 c.p.m. ml⁻¹. After hybridization, the filters were washed three times in 250 ml of $2 \times$ SSC plus 0.1% SDS at room temperature, and twice in 250 ml of $2 \times$ SSC plus 0.1% SDS at 50°C for 15–20 min each. The filters were air-dried and exposed to Kodak XAR-5 film at -70°C using an intensifying screen. Positive plaques were purified by one additional cycle of screening. Single plaques were isolated and amplified through *Escherichia coli* KM392, followed by a medium-scale phage preparation as described previously²⁷. Phage DNA was prepared according to standard procedures²⁷, digested with *Eco*RI and the generated insert was isolated from 0.8% agarose slab gel by adsorption to activated glass²⁸. Plasmid SP65 (ref. 20) was linearized with *Eco*RI and the ends were treated with calf intestinal phosphatase to make it suitable as cloning vehicle²⁹. Vector and cDNA insert were mixed in a 1:2 molar ratio, followed by ligation in the presence of 1 mM ATP and 10 units of T4 ligase for 13–16 h at 16°C. Part of the ligation mixture was used to transform *E. coli* MC1061 (ref. 30) to ampicillin resistance. Recombinants were isolated, identified and grown in large cultures using standard techniques²⁷. Plasmids were purified through CsCl gradients²⁷. Purified cDNA insert was digested with *Xba*I, *Hinf*I, *Taq*I and *Dde*I, and the generated ends were labelled either at their 5' ends with T4 polynucleotide kinase and [γ -³²P]ATP or at their 3' ends with DNA polymerase I (Klenow fragment) and [α -³²P]dNTPs. The labelled fragments were separated by electrophoresis in a 5% polyacrylamide gel and the individual fragments were isolated and purified²¹. These fragments were subsequently treated with a buffer containing 30% (v/v) dimethyl sulphoxide and 1 mM EDTA for 3 min at 90°C and run through a strand-separating gel²¹. Single strands were isolated, purified and subjected to chemical degradation as described by Maxam and Gilbert²¹.



segment (residue 80–106) and a cytoplasmic domain (residues 107–152). A high level of amino-acid sequence conservation is apparent both in the transmembrane (70.4%) and intracellular (77.2%) domains of murine and human T3 δ -chain. Much less homology was found in the extracellular domains (58.2%).

The locations of the five cysteines (residues 16, 52, 72, 75 and 103) and two attachment sites for Asn-linked oligosaccharides (residues 17 and 53) are conserved (Figs 2, 3). The murine T3 δ -chain, however, contains a third site for Asn-linked glycans at residue 34. The existence of this third consensus sequence (Asn 34–Lys 35–Thr36) for N-linked glycosylation in the murine T3 δ -chain suggests that the mature glycosylated murine T3 δ -chain has a higher relative molecular mass than its human

equivalent. The attachment of three oligosaccharide moieties to a relatively short extracellular domain (79 residues) might decrease the immunogenicity of the murine T3 δ ; this could explain why attempts to prepare rat monoclonal reagents specific for murine T3 have been unsuccessful. The amino-acid sequences around Asn 34 are very different from those in the human polypeptide chain. This suggests that murine anti-T3 monoclonal antibodies recognize a determinant which is located between residues 20 and 39 in the human T3 δ -chain. As this area of the molecule carries an oligosaccharide in the murine chain, this putative determinant is probably also located on the surface of the human protein. The existence of an easily accessible determinant on the surface of the human molecule would

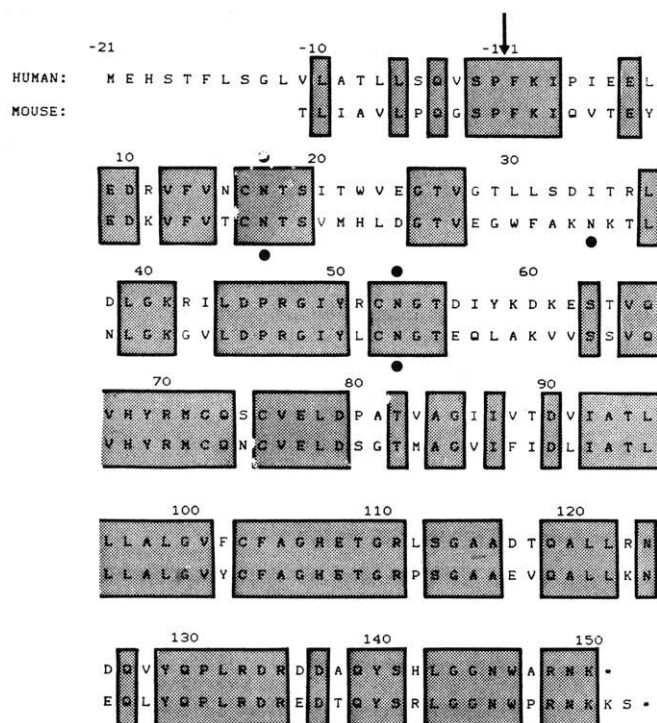


Fig. 2 Comparison of the amino-acid sequences of murine and human T3 δ -chain. The vertical arrow above the sequence indicates the first amino acid of the mature protein. Identical amino acids are boxed. The potential attachment sites for asparagine-linked glycosylation are indicated by solid circles.

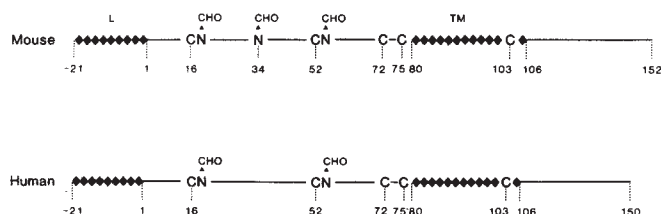


Fig. 3 General features of the organization of the murine and human T3 δ -chain. L represents the leader peptide sequence, which is presumably 21 amino acids long in murine T3 δ -chain, and TM represents the transmembrane domain. C, cysteine residues; CHO, N-linked glycosylation sites.

explain why murine anti-T3 monoclonal antibodies have been prepared with a relatively high frequency.

Of interest is the conservation of the aspartic acid residue at position 90 in the transmembrane domain (Fig. 2). This amino acid could be involved in a salt bridge with the lysine or arginine residues present in the transmembrane segment of the T-cell receptor α - and β -chains (Fig. 4)^{14,15,25,26}. The hydrophobic environment may enhance the stabilizing effect of this salt bridge on the protein/protein interactions between the T3 δ -chain and the α - and β -chain of the T-cell receptor. Such interaction depends, of course, on the precise positioning of the charged residues.

Thus, we have identified the murine homologue of the T3 δ -chain. This finding, together with the similarity in structure of the T-cell receptor α - and β -chains, predicts that murine γ - and ϵ -chains will be found also. The availability of a cDNA clone coding for the murine T3 δ -chain will allow the generation of anti-peptide antibodies, which will be useful in studies of the T3/T-cell receptor complex on the surface of antigen-specific mouse T cells. Furthermore, these antibodies, together with the DNA probes, will allow us to correlate transcriptional activity with cell-surface expression of these cell-surface structures within the thymus. This knowledge should further our understanding of the regulation of the function of the T3/T-cell

Mouse T3 δ -chain SGTMAGVIFIDLIATLLALLGVYCFAG

Mouse α NLSVMGLRIILLKLVAGFNLLMTL

Mouse β ILLGKATLYAVLVSGGLVLMAMV

Fig. 4 Amino-acid sequences of the transmembrane regions of murine T3 δ -chain and T-cell receptor α - and β -chains aligned at their C-termini. Charged amino-acid residues are indicated.

receptor complex during thymic differentiation and embryonic development.

We thank Drs Antonella Spadaro, Maryellen Ruvolo, Kut-nie Tan, Linda Clayton and Allan Maxam for making the λ gt10 library available, and Sharon McGrath for typing the manuscript. This work was supported by NIH grants AI-15066 and AI-17651. P.v.d.E. is a recipient of a Niels Stensen Stichting fellowship. C.T. is a Scholar of the Leukemia Society of America.

Received 11 December 1984; accepted 7 January 1985.

1. Borst, J., Prendiville, M.-A. & Terhorst, C. *J. Immun.* **128**, 1560-1565 (1982).
2. Borst, J., Alexander, S., Elder, J. & Terhorst, C. *J. Biol. Chem.* **258**, 5135-5141 (1983).
3. Borst, J., Prendiville, M. A. & Terhorst, C. *Eur. J. Immun.* **13**, 576-580 (1983).
4. Borst, J. *et al. Nature* **312**, 455-458 (1984).
5. Kanellopoulos, J. M., Wigglesworth, N. M., Owen, M. J. & Crumpton, M. J. *EMBO J.* **2**, 1807-1814 (1983).
6. Meuer, S. C., Acuto, O., Hercend, T., Schlossman, S. F. & Reinherz, E. L. *A. Rev. Immun.* **2**, 23-50 (1984).
7. Oettingen, H. C., Kappler, J., Tax, W. J. M. & Terhorst, C. *J. Biol. Chem.* **259**, 12039-12048 (1984).
8. Van Wauwe, J. P., DeMey, J. R. & Goossens, J. G. *J. Immun.* **124**, 2708-2713 (1980).
9. Terhorst, C. *Receptors Recognition* **16**, 131-162 (1984).
10. Oettingen, H. C., Terhorst, C., Cantley, L. C. & Rosoff, P. *Cell* (in the press).
11. Samelson, L. E., Germain, R. N. & Schwartz, R. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972-6976 (1983).
12. Kappler, J. *et al. Cell* **35**, 295-302 (1983).
13. Acuto, O. *et al. J. exp. Med.* **158**, 1368-1373 (1983).
14. Yanagi, Y. *et al. Nature* **308**, 145-148 (1984).
15. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153-157 (1984).
16. van den Elsen, P. *et al. Nature* **312**, 392-397 (1984).
17. Glimcher, L. H. & Shevach, E. M. *J. exp. Med.* **156**, 640-645 (1982).
18. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
19. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning: A Practical Approach* (ed. Glover, D.) (IRL, Oxford, 1984).
20. Green, M. R., Maniatis, T. & Melton, D. A. *Cell* **32**, 681-694 (1983).
21. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
22. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726-730 (1982).
23. Barker, W. C. *et al. Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Fdn, Washington, DC, 1984).
24. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
25. Chien, Y. *et al. Nature* **312**, 31-35 (1984).
26. Saito, H. *et al. Nature* **312**, 36-40 (1984).
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
28. Vogelstein, B. & Gillespie, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 615-619 (1979).
29. van den Elsen, P., Houweling, A. & Van der Eb, A. *Virology* **128**, 377-390 (1983).
30. Casadaban, M. J. & Cohen, S. N. *J. molec. Biol.* **138**, 179-207 (1980).

An unusual type of V-J joining diversifies the primary repertoire of mouse λ 1 light chains

Klaus Karjalainen* & Christopher Coleclough†

* Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland

† Roche Institute of Molecular Biology, Nutley, New Jersey 07110, USA

It is well established that B lymphocytes can achieve an almost unlimited antibody repertoire by using combinations of at least three different basic mechanisms¹. First, the mammalian genome has multiple, distinct germ-line variable (V), joining (J) and diversity (D) gene segments which can presumably combine randomly to give V-D-J joins in the heavy (H)-chain loci and V-J joins in the light (L)-chain loci during the formation of functional antibody genes. Second, the actual joining points between any two

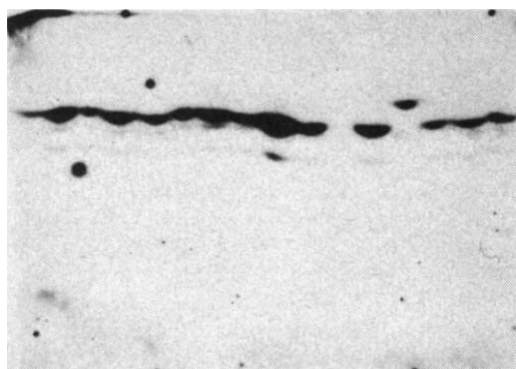


Fig. 1 IEF analysis, which shows one variant $\lambda 1$ light chain. To obtain the $\lambda 1$ -chain-containing supernatants, normal spleen cells from (BALB/c $\times$ NZB) F_1 mice were cultured in the presence of $50 \mu\text{g ml}^{-1}$ LPS (*Salmonella typhosa* 0901, Difco) in RPMI medium supplemented with 20% fetal bovine serum, 5×10^{-5} M 2-mercaptoethanol and $50 \mu\text{g ml}^{-1}$ gentamycin. The initial cell concentrations were $5 \times 10^5 \text{ ml}^{-1}$ and $3 \times 10^4 \text{ ml}^{-1}$ for day 3 and day 8 cultures respectively. Stimulated cells were fused with Sp2-0 plasmacytoma cells using standard procedures⁶. When 10^8 fused cells were plated into 20 96-well microtitre plates, multiple clones grew in individual wells. Cells from those wells that showed a positive reaction with a monoclonal rat anti-mouse $\lambda 1$ antibody (4/1-101)³ were subcloned, and the supernatants from these $\lambda 1$ -positive subclones were subjected to IEF analysis. The supernatants were mildly reduced with dithiothreitol (10 mM, 15 min at room temperature) followed by alkylation with $2 \times$ molar excess of iodoacetamide (30 min at room temperature). Just before loading, a fourfold volume of IEF sample buffer (9.5 M urea and 2% ampholine pH 3.5-10) was added to the supernatants. Usually $\sim 40 \mu\text{l}$ of solution was applied per lane to 0.8-mm-thick 5% polyacrylamide slab gels (model V-16-2; BRL) containing 9.2 M urea and 2% ampholine pH 3.5-10. After overnight focusing at 200 V, the proteins were electrophoretically transferred to nitrocellulose filters using Western blot conditions⁷, then the nitrocellulose filters were incubated in blocking solution (1% casein, 50 mM Tris-HCl pH 8.0, 150 mM NaCl) for 1 h at room temperature. Finally, $\lambda 1$ light chains were detected using iodinated anti- $\lambda 1$ monoclonal antibody, with subsequent autoradiography of the washed filters.

combining gene segments can vary considerably, increasing the germline diversity. Further variability is generated in the heavy-chain locus by *de novo* addition of extra nucleotides between the combining gene segments. Finally, somatic mutations independent from joining events can accumulate in V regions during the lifetime of B lymphocytes. Here we report that when V and J regions join in the formation of functional $\lambda 1$ light-chain genes, 'lethal' out-of-frame joins can be compensated for by the deletion of nucleotides several bases upstream of the actual joining points; this generates small stretches of nucleotides in a new frame between the deletion and the V-J joining point, thus creating additional diversity in the third hypervariable regions of the mouse $\lambda 1$ light chains.

In order to be able to study the mechanisms which generate diversity in the variable regions of mouse antibodies, we decided to minimize the complexity of the experimental system in two ways. First, we concentrated on $\lambda 1$ light chains because the inherited repertoire of $\lambda 1$ genes contains only one V region whose sequence is known¹; thus, any different sequences found must result from somatic mutation. Second, we studied only virgin primary B cells and not, for example, immune cells, so that the frequency of variants observed would reflect as closely as possible the true frequency of somatic variation. To do this we stimulated normal spleen cells from (BALB/c $\times$ NZB) F_1 mice with bacterial lipopolysaccharide (LPS) which triggers the proliferation of virgin mouse spleen B cells². Stimulated cells were then immortalized by fusion with Sp2-0 plasmacytoma cells after 3 or 8 days of *in vitro* culture.

We obtained 101 and 113 $\lambda 1$ -secreting clones, as defined by their reaction with rat monoclonal anti-mouse $\lambda 1$ antibody

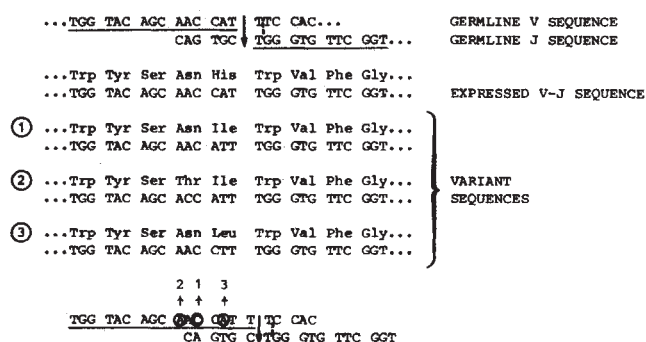


Fig. 2 The V-J join most frequently found is shown at the top, with the coding sequences of the V and J gene segments prior to rearrangement underlined. This 'germline' $\lambda 1$ join can be compared with those of the variants, whose sequences were determined. All three variants have an extra T residue inserted at the V-J junction, and this is compensated for by the deletion of a single residue (circled nucleotides at the bottom) in the adjacent V sequence.

Methods. A double-stranded *SacI*/*KpnI* restriction fragment (75 base pairs (bp) long) corresponding to the 5' portion of the $\lambda 1$ constant-region gene was excised from the complementary DNA clone⁸ pABA1-15 (obtained from Dr A. Bothwell) and used as a primer. Total cytoplasmic RNA (50 ng) and the 5' end-labelled primer (200 ng) were mixed in a 100- μl volume of annealing buffer (80% dimethylformamide, 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA) and incubated at 47.5 °C for 90 min. Annealing was stopped by diluting the mixture fourfold with 0.3 M sodium acetate and by precipitating nucleic acids with ethanol. The resultant pellet was dissolved in standard reverse transcription buffer and primer extension carried out using established conditions for reverse transcription⁹. Full-length cDNA was isolated on denaturing 5% polyacrylamide gels, and the sequence of the gel-purified material was determined according to the method of Maxam and Gilbert¹⁰.

4/1-101, (ref. 3), from day 3 and day 8 fusions, respectively. Ninety-five per cent of the day 3 but only 40% of the day 8 clones were IgM producers, showing that our culture conditions, like the limiting dilution cultures described by others⁴, allowed the immunoglobulin switch to occur. To detect variants, we used isoelectric focusing (IEF) analysis as a primary screen; although this method underestimates the DNA sequence variability, the IEF method that we developed made it easy to analyse hundreds of supernatants (see Fig. 1 legend). We found four IEF variants among 213 $\lambda 1$ chains studied ($\sim 2\%$). (Figure 1 shows an example of a primary screen that revealed one variant.) All four variants had similar IEF patterns; these were one charge unit more acidic than the wild-type pattern. The variants seemed to represent a random subpopulation of the whole sample of 213 $\lambda 1$ -secreting clones: two originated from the day 3 fusion, being of IgM type, and of the two day 8 clones, one had switched to IgG.

We determined the DNA sequences of three of the variants and of one wild type by using the primer extension technique (see Fig. 2 legend). It was immediately apparent from the sequence data that the variants were the products of a novel kind of joining mode. All of them had an extra T nucleotide at the V-J junction compared with the most frequent V-J join (Fig. 1). To maintain the correct reading frame, a compensatory deletion of one base had occurred 2, 4 and 6 nucleotides upstream of the 'inserted' T residue; as a result, only a small stretch of nucleotides was out-of-frame between the deletion and the insertion. In the most extreme case (variant 2) the DNA sequence encodes two new amino acids, Thr-Ile, at positions 95 and 96 compared with the wild-type sequence (Asn-His). The switch from the basic amino acid His to the neutral Ile or Leu (His is positively charged at the isoelectric point of the $\lambda 1$ light chain, $pI \sim 4.5$) in these variants explains the acidic shift in their IEF pattern. No other mutations were found when about 80% of the V regions of the variants were sequenced (data not shown).

Explanations of the V-J joining mechanism must be consistent with the above findings. The extra T at the V-J junction is probably derived from the region just 3' of the $\lambda 1$ V region (see Fig. 2), that is, the recombination point is moved by one nucleotide. To explain the deletions upstream of the actual join, we speculate that V-J recombinations have error-prone sealing and repair activities. We do not know how frequently small deletions accompany the joining process nor do we know whether the deletions can occur also at the 3' side of the junction because we have analysed only a selected subpopulation of $\lambda 1$ chains. Certainly, the one-base deletion does not always occur when an extra T is added at the V-J join. An aberrantly rearranged $\lambda 1$ light-chain gene having an 'extra T' but lacking the one-base deletion has been described⁵.

Clearly, more data are needed to establish whether the mechanism described here is an important source of diversity in antigen receptors which are formed by the joining of two separate gene segments.

We thank Carolyn Tschachtli for preparation of the manuscript, and Drs David Nemazee and Siegfried Weiss for critical reading and comments. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co., Basel, Switzerland.

Received 29 October 1984; accepted 28 January 1985.

1. Tonegawa, S. *Nature* **302**, 575-581 (1983).
2. Freitas, A. A. & Coutinho, A. *J. exp. Med.* **154**, 994-999 (1981).
3. Takemori, T. & Rajewsky, K. *Eur. J. Immun.* **11**, 618-625 (1981).
4. Andersson, J., Coutinho, A. & Melchers, F. *J. exp. Med.* **147**, 1744-1754 (1978).
5. Hozumi, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 7019-7023 (1981).
6. Köhler, G. & Milstein, C. *Nature* **256**, 495-496 (1975).
7. Towbin, H., Staehlin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
8. Bothwell, A. L. M. *et al. Nature* **298**, 380-382 (1982).
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, A Laboratory Manual* (1982).
10. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

Induction of *c-fos* during myelomonocytic differentiation and macrophage proliferation

Rolf Müller, Tom Curran*, Dagmar Müller & Larry Guilbert

European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, FRG

* Department of Molecular Genetics, Roche Research Center, Hoffmann-La Roche Inc., Nutley, New Jersey 07110, USA

Previous studies have suggested a role for *c-fos* in cellular differentiation in fetal membranes¹⁻³, haematopoietic cells^{4,5} and teratocarcinoma stem cells⁶. In other cell types, such as fibroblasts, *c-fos* expression is normally very low, but is rapidly induced by peptide growth factors, implicating *c-fos* in growth control mechanisms⁷⁻¹⁰. Here, we show that the TPA (12-*O*-tetradecanoylphorbol-13-acetate)-induced macrophage-like differentiation of HL60 human promyelocytic precursor cells^{11,12} is accompanied by the induction of both *c-fos* mRNA and protein within 15 min after treatment, suggesting a functional role for *c-fos* in this differentiation system. In quiescent terminally differentiated macrophages, expression of *c-fos* is inducible by the macrophage-specific growth factor colony-stimulating factor-1 (CSF-1)¹³. The kinetics of *c-fos* induction, however, are entirely different from those in growth factor-stimulated fibroblasts, supporting the view that the *c-fos* gene product may serve different functions in different cell types.

The human promyelocytic leukaemia cell line HL60 (ref. 11) can be induced by the tumour promoter TPA to differentiate into macrophage-like cells¹². Previously, it has been shown that

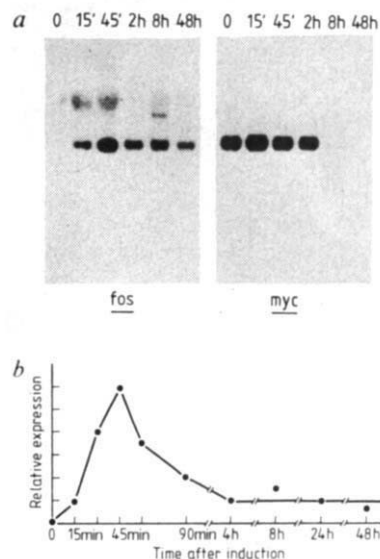


Fig. 1 Expression of *c-fos* and *c-myc* in TPA-induced differentiating HL60 cells. *a*, Northern blot analysis of total RNA (16 µg per lane) isolated at the indicated times after induction by 5×10^{-7} M TPA. Methods as described elsewhere⁸. Hybridization to *v-fos* and mouse *c-myc*-specific probes was performed as described⁸. *b*, Relative expression of *c-fos* plotted against time after TPA treatment. Quantitative evaluation of the Northern blot autoradiograms (similar to that shown in *a*) was done by comparing band intensities.

this differentiation is associated with a drastic reduction in the level of expression of *c-myc* mRNA^{5,14} and an increase in the level of *c-fos* mRNA⁵ at a time when differentiation is almost complete (that is, 48 h after induction). To investigate a possible involvement of *c-fos* in the induction of differentiation, the time course of *c-fos* expression following TPA treatment of HL60 cells was followed in detail. Figure 1*a,b* shows that *c-fos* mRNA was not detectable in normal HL60 cells, but treatment with 5×10^{-7} M TPA induced *c-fos* expression within 15 min. Maximum expression was observed 30-45 min after addition of TPA; thereafter, the level of *c-fos* mRNA decreased 3-5-fold and remained at this level for the following 48 h. In contrast, expression of *c-myc* mRNA did not change until 4 h after TPA addition, when expression fell by more than 20-fold (Fig. 1*a*).

Differentiation of HL60 cells can also be induced by compounds other than TPA. A vitamin D₃ dihydroxy derivative, 1,25(OH)₂D₃, induces monocytic differentiation¹⁵, whereas treatment with retinoic acid gives rise to granulocytic cells¹⁶. In both cases, however, rapid induction of *c-fos* mRNA was not detectable (data not shown). In another study⁴, it was reported that granulocyte-colony stimulating factor (G-CSF)-induced mouse myelomonocytic precursor cells express *c-fos* only at late stages of differentiation (that is, >2 days after induction). Taken together, these observations show that *c-fos* expression at an early stage is not an obligatory step in the differentiation of myelomonocytic or granulocytic cells. Rapid induction of *c-fos* expression seems to be specific for the TPA-induced pathway of monocytic differentiation.

Because *c-fos* protein has not yet been detected in human cells, we investigated whether the rapid induction of *c-fos* mRNA expression is associated with induction of *c-fos* protein, by immunoprecipitation of ³⁵S-labelled proteins from HL60 cells induced by TPA for various times. We used an antiserum directed against a synthetic peptide (designated 'M') common to *v-fos* and both mouse and human *c-fos* sequences¹⁷. It has been shown previously that this antiserum detects modified *c-fos* protein of relative molecular mass *M*_r 57,000-62,000 (57-62 K) which is complexed with a cellular protein, p39 (refs 8, 17-19;

Fig. 2 Immunoprecipitation of *c-fos*-related protein from HL60 cells induced to differentiate by 50×10^{-8} M TPA (a) or 10^{-7} M $1,25(\text{OH})_2\text{D}_3$ (b) for different times. C, Control serum (*v-fos* peptide-specific antiserum; V in ref. 17); M, *fos* peptide-specific antiserum¹⁷. For comparison, immunoprecipitation of *c-fos* protein from FCS-stimulated NIH 3T3 cells is also shown (a). With HL60 cells, immunoprecipitation was performed on lysates of ³⁵S-methionine-labelled cells denatured by boiling for 5 min in 0.5% SDS¹⁷, thus no p39 is co-precipitated in these lanes. The X protein was not found in non-denatured lysates, and its level was subject to considerable experimental variation in denatured cell lysates, suggesting that it may be a cross-reacting protein that shares only very restricted antigenicity with the *c-fos* gene product.

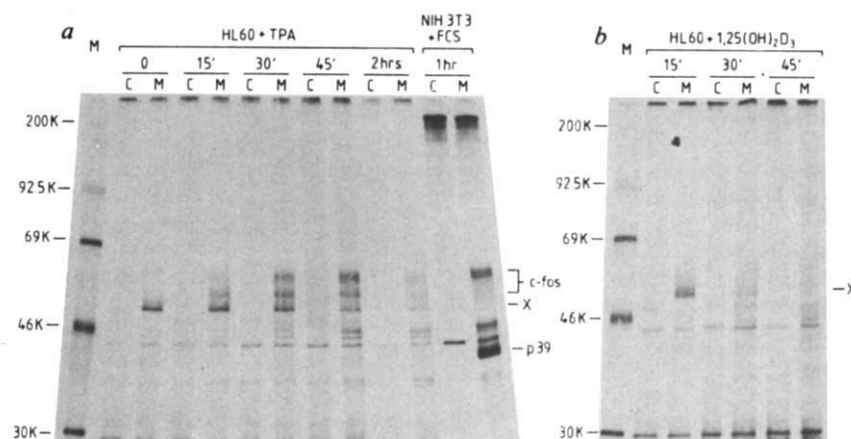


Fig. 2a). Protein of very similar M_r was detected in the HL60 cells as early as 15 min after TPA induction. The levels of *c-fos* protein were maximal at 30–45 min, then decreased (Fig. 2a). The cellular p39 protein was not detected in this experiment because immunoprecipitation was performed on cell lysates denatured by boiling in SDS. Very similar results were obtained with a *fos*-specific tumour-bearing rat (TBR) serum^{19,20} (data not shown). In agreement with RNA analyses, expression of *c-fos* protein was not detectable in $1,25(\text{OH})_2\text{D}_3$ -induced HL60 cells (Fig. 2b). Another group of proteins in the 45 K range increased concomitantly with *c-fos* in both TPA-induced HL60 cells and—as described previously⁸—in fetal calf serum (FCS)-stimulated NIH 3T3 cells. These proteins may be encoded by a *c-fos* related gene, such as *r-fos*, which, in growth factor-stimulated fibroblasts, is co-induced with *c-fos*¹⁰. Coincidentally, the *fos*-specific antiserum was raised against a peptide that is also contained in the *r-fos* protein^{10,17}. A more precise definition of these proteins awaits peptide mapping analyses. Interestingly, a protein of $M_r \sim 55$ K (designated 'X' in Fig. 2) was detectable in denatured lysates of uninduced HL60 cells and disappeared rapidly after induction of differentiation by either TPA or $1,25(\text{OH})_2\text{D}_3$ (Fig. 2a,b). As *c-fos* RNA is not detectable in uninduced HL60 cells (Fig. 1a), the X protein apparently cannot be encoded by *c-fos*. Although this protein was also detected by the *fos*-specific TBR serum (data not shown) its nature remains obscure.

In fibroblasts, *c-fos* expression is rapidly inducible by growth factors and is therefore thought to be associated with growth regulation. Although macrophages are terminally differentiated cells, they still retain the ability to proliferate, unlike most other cell types. Therefore, it was of interest to investigate whether *c-fos* expression in mature macrophages might be correlated with growth. When primary cultures of bone marrow-derived macrophages were deprived of the macrophage-specific growth factor CSF-1 for 18 h (Fig. 3a) or kept at low serum concentrations for 24 h (0.5% FCS; data not shown), proliferation ceased and the cells became quiescent. This was accompanied by a marked reduction in levels of both *c-fos* and *c-myc* mRNA (Fig. 3b, and data not shown), suggesting that expression of these proto-oncogenes depends on the presence of growth factors (or on cellular proliferation). When CSF-1-deprived macrophages in the G_0 phase of the cell cycle were stimulated by addition of the growth factor, induction of *c-myc* was observed after 1 h, reaching a maximum at 2–4 h. Activation of *c-myc* was followed by an ~ 10 -fold induction of *c-fos* mRNA levels, which remained high for the following 8 h (Fig. 3b). Regulation of *c-fos* expression in macrophages differs substantially from that in fibroblasts: (1) in contrast to fibroblasts⁸, a high level of *c-fos* expression is evident in normally growing, asynchronous macrophages, and thus appears to be part of the normal cell cycle in these cells; (2) induction of *c-fos* by CSF-1 in macrophages occurs much later than in growth factor-stimulated

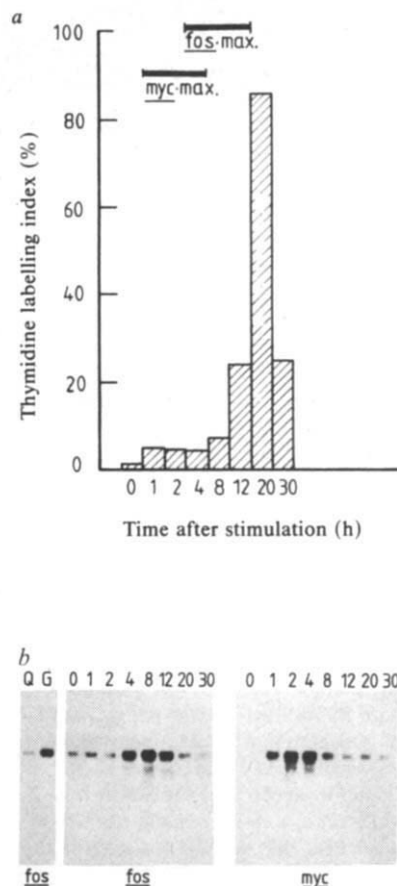


Fig. 3 a, Thymidine labelling index at the indicated times after CSF-1 induction of quiescent macrophages. Horizontal bars indicate the periods of maximal *c-fos* and *c-myc* expression, respectively, as observed in the experiment shown in b. b, Expression of *c-fos* and *c-myc* in quiescent versus growing or stimulated macrophages. Northern blot analysis of RNA from normally growing (G) versus CSF-1-deprived, quiescent (Q) mouse macrophages, and from quiescent macrophages stimulated with CSF-1 for the indicated times (0, 1, 2, 4, 8, 12, 20 and 30 h).

Methods. Primary cultures of bone marrow-derived macrophages from C3H/HeJ mice were established and grown as described previously¹³. Cultures were used at day 7 after establishment. At this time the population consisted of $>98\%$ macrophages. Quiescent macrophages were obtained by maintaining the culture for 18 h in CSF-1-free medium supplemented with 15% FCS. Stimulation was carried out by adding 3,000 units of partially purified CSF-1 (stage 1) per ml of medium. Purification of CSF-1 from mouse L-cell supernatant has been described previously²². One dish (175 cm^2) of macrophages was used for each time point.

fibroblasts, where it precedes activation of *c-myc*; (3) induction of *c-fos* in macrophages lasts considerably longer than in fibroblasts. While in stimulated fibroblasts *c-fos* induction appears to occur during the transition from the G_0 to G_1 phase of the cell cycle, expression in macrophages is observed during G_1 until mid-S phase. These findings further substantiate our hypothesis^{6,8} that the *c-fos* gene product serves different functions in fibroblasts and macrophages.

TPA induces cell proliferation in fibroblasts²¹ and differentiation in HL60 cells¹². In both cases, rapid induction of *c-fos* mRNA and protein was observed (Figs 1, 2; ref. 7 and R. Bravo, J. Burckhardt, T.C. and R.M., unpublished observation). It is possible that the initial biochemical events induced by TPA are

the same regardless of whether the ultimate consequence is differentiation or proliferation. Such initial events, in which *fos* may have a role, include the transduction of TPA-induced signals received at the cell membrane within the nucleus. In mature myelomonocytic cells, *c-fos* may be involved in biochemical pathways controlling proliferation and/or securing the survival of macrophages. In this context it is interesting that *c-fos* is induced by the growth factor CSF-1, which has been shown to be required for both growth and survival of macrophages¹³.

We thank R. Bravo for valuable discussions, M. R. Uskokovic for 1,25(OH)₂-D₃, C. Brady for technical assistance, and B. Blanasch for help in preparation of the manuscript.

Received 3 December 1984; accepted 28 January 1985.

1. Müller, R., Slamon, D. J., Tremblay, J. M., Cline, M. J. & Verma, I. M. *Nature* **299**, 640-644 (1982).
2. Müller, R., Verma, I. M. & Adamson, E. D. *EMBO J.* **2**, 679-684 (1983).
3. Müller, R., Tremblay, J. M., Adamson, E. D. & Verma, I. M. *Nature* **304**, 454-456 (1983).
4. Gonda, T. J. & Metcalf, D. *Nature* **310**, 249-251 (1984).
5. Müller, R., Müller, D. & Guilbert, L. *EMBO J.* **3**, 1887-1890 (1984).
6. Müller, R. & Wagner, E. F. *Nature* **311**, 438-442 (1984).
7. Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
8. Müller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716-720 (1984).
9. Kruijer, W., Cooper, J., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
10. Cochran, B. H., Zullo, J., Verma, I. M. & Stiles, C. D. *Science* **226**, 1080-1082 (1984).

11. Collins, S. J., Gallo, R. C. & Gallagher, R. W. *Nature* **270**, 347-349 (1977).
12. Rovera, G., O'Brien, T. G. & Diamond, L. *Science* **204**, 868-870 (1979).
13. Tushinsky, R. J. *et al. Cell* **28**, 71-78 (1982).
14. Westin, E. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2490-2494 (1982).
15. McCarthy, D. M. *et al. Leukemia Res.* **7**, 51-55 (1983).
16. Reitsma, P. H. *et al. Nature* **306**, 492-494 (1983).
17. Curran, T., Van Beveren, C., Ling, N. & Verma, I. M. *Molec. cell Biol.* **5** (in the press).
18. Curran, T. & Teich, N. M. *Virology* **116**, 221-235 (1982).
19. Curran, T., Miller, A. D., Zokas, L. & Verma, I. M. *Cell* **36**, 259-268 (1984).
20. Curran, T. & Teich, N. M. *J. Virol.* **42**, 114-122 (1982).
21. Sivak, H. J. *cell. Physiol.* **80**, 167-174 (1972).
22. Stanley, E. R. & Guilbert, L. Y. *J. immun. Meth.* **42**, 253-284 (1981).

Increased neurovirulence associated with a single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome

D. M. A. Evans*, G. Dunn*, P. D. Minor*,
G. C. Schild*, A. J. Cann†, G. Stanway†,
J. W. Almond†, K. Currey‡ & J. V. Maizel Jr‡

* National Institute for Biological Standards and Control, Holly Hill, Hampstead, London NW3 6RB, UK

† University of Leicester, University Road, Leicester LE1 7RH, UK

‡ Laboratory of Molecular Genetics, National Institutes of Health, Bethesda, Maryland 20205, USA

Most of the small number of cases of poliomyelitis which occur in countries where Sabin's attenuated poliovirus vaccines are used are temporally associated with administration of vaccine and involve polioviruses of types 2 and 3 (ref. 1). Recent studies have provided convincing evidence that the Sabin type 2 and 3 viruses themselves may revert to a neurovirulent phenotype on passage in man²⁻⁶. We report here that a point mutation in the 5' noncoding region of the genome of the poliovirus type 3 vaccine consistently reverts to wild type in strains isolated from cases of vaccine-associated poliomyelitis. Virus with this change is rapidly selected on passage through the human gastrointestinal tract. The change is associated with a demonstrable increase in the neurovirulence of the virus.

The complete nucleotide sequences of the genomes of the neurovirulent progenitor of the Sabin type 3 vaccine virus, P3/Leon/37, the Sabin vaccine virus itself, P3/Leon/12a,b, and a virulent revertant strain from the brain of a fatal case of vaccine-associated poliomyelitis, P3/119 (ref. 7), have been determined^{4,7-9}. The small number of sequence differences between these strains^{4,7} suggested that the following mutations could be involved in attenuation and reversion: (1) a single nucleotide change at position 472 in the 5' noncoding region of the genome; (2) nucleotide changes resulting in amino-acid changes in the capsid proteins; and (3) a single change at the end of the 3' noncoding region. Only the nucleotide at position 472 correlated with the neurovirulent phenotype, being cytidine in P3/Leon/37 and P3/119, and uridine in the vaccine strain.

Table 1 Base at position 472, time of isolation, neurovirulence and temperature sensitivity of Sabin type 3 vaccine-derived strains of poliovirus

Virus	Base at position 472	Time of isolation after vaccination	Mean histological lesion score	rct marker test*
Sabin vaccine	U		0.36	>5.5 (rct ⁻)
DM1	U	24 h	ND	ND
DM2	U	31 h	1.58	6.13 (rct ⁻)
DM3	U/C	35 h	ND	ND
DM4†	C	47 h	2.48	5.71 (rct ⁻)
DM38	C	18 days	ND	ND
DM119	C	3-4 weeks	3.34	0.25 (rct ⁺)

Mean histological lesion scores were determined using the standard WHO neurovirulence test¹⁴. The range of mean histological lesion scores of a type III attenuated reference strain in eight tests carried out over the past 2 yr is 0.43-0.79. ND, not determined.

* The rct marker (replicative capacity at temperature -40°C) test shows the logarithmic difference between virus titres grown at 35°C and at 40°C¹⁵. Viruses showing reductions of log 5.0 or more in titre are considered to be temperature-sensitive (rct⁻).

† The sequences of the structural coding region and the 5' noncoding region of DM4 were determined as described previously¹⁰. The extreme 3' end of the RNA was sequenced by synthesizing complete complementary DNA copies of the viral RNA, annealing a synthetic oligonucleotide primer to its 5' end and sequencing by the dideoxy method using the Klenow fragment of *Escherichia coli* DNA polymerase I¹³.

We therefore examined the frequency with which the mutation to cytidine at base 472 occurred in viruses from vaccine-associated cases of poliomyelitis, and assessed its effect on the phenotype of the virus.

Twelve strains of poliovirus type 3 from a collaborative study of vaccine-associated poliomyelitis performed by the World Health Organization (WHO)⁹ were examined in addition to P3/119. All the strains had been shown to be genetically related to the Sabin vaccine virus by T1 oligonucleotide mapping^{2,3,6}, and six had been found to be highly neurovirulent⁹. In all 12 strains, the nucleotide at position 472 of the genome was cytidine, as shown by primer extension sequencing¹⁰ using a synthetic primer which annealed to the viral RNA from position 510 to 522 (data not shown). In contrast, the nucleotide at this position in two batches of vaccine was uridine, confirming our previous findings^{4,8,9} and those of others¹¹. The consistency with which cytidine was detected at position 472 amongst the vaccine-derived revertant viruses suggested that this nucleotide might have an effect on the virulence of the virus.

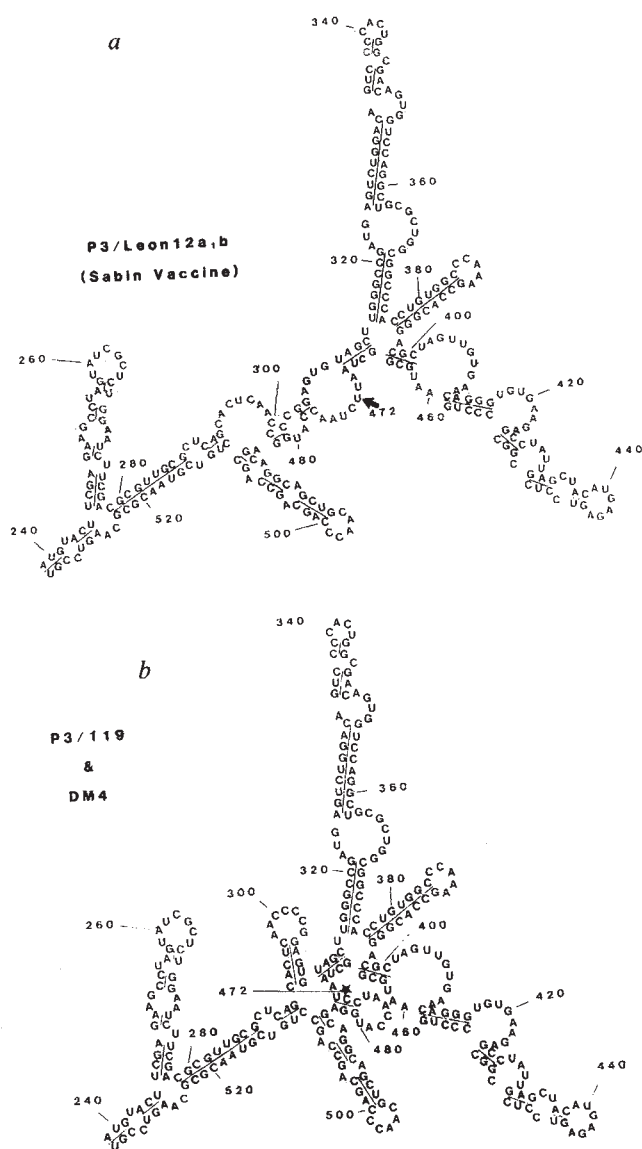


Fig. 1 Predicted secondary structure of RNA (bases 240-540, approximately) from type 3 Sabin poliovirus vaccine strain (a) and type 3 revertant strain P3/119 (b) based on the improved RNZ folding algorithm of Jacobson *et al.*¹⁸. We used a recent display algorithm which allows easy visual comparison of similar sequences¹⁹.

It is known that Sabin type 3 vaccine virus increases in neurovirulence on passage in the human gut¹². Therefore, we attempted to determine whether this loss of attenuation is associated with the substitution at position 472. After oral administration of the trivalent Sabin vaccine to a 4-month-old infant, we collected every faecal sample and attempted to isolate the virus from each. The genomes of several of the successive type 3 isolates were sequenced through position 472 (Table 1). DM1 and DM2, isolates from the first two faecal samples obtained 24 and 31 h after vaccination respectively, possessed U at position 472, as in the vaccine virus. The third isolate, DM3, from a faecal sample collected at 35 h, contained a mixture of viruses with either U or C at position 472. The fourth isolate (DM4), from a sample collected just 47 h after vaccination, and all subsequent isolates sequenced appeared to be homogeneous for C at position 472. These findings suggest strongly that replication of the vaccine virus in the human gut rapidly selects for viruses containing C at position 472 of the genome. No other changes, including those with significance for reversion^{4,13}, were detected when the entire structural coding region and the 3' and 5' noncoding regions of DM4 were sequenced (Table 1).

To assess the significance of the mutation at position 472 for neurovirulence, DM2 and DM4 were assayed in a standard WHO neurovirulence test¹⁴. DM4 was found to be significantly more neurovirulent than the vaccine strain with respect to both the histological lesions in the central nervous system (CNS) (Table 1) and the paralysis induced (data not shown), although it was not as neurovirulent as the revertant P3/119 (Table 1). The second isolate, DM2, had a higher lesion score than the Sabin vaccine strain. Our interpretation of this finding is that DM2 probably contained a small proportion of virus with C at position 472 which was undetectable by the sequencing methods used.

DM4 was found to be temperature-sensitive (*rct*⁻) (Table 1), as is the Sabin vaccine virus¹⁵. In contrast, the revertant P3/119 was *rct*⁺, as were all other revertants tested⁹. Therefore, it seems that reversion at position 472 alone does not account for the *rct*⁺ phenotype of revertant strains.

As the extremely rapid selection of virus having cytidine at position 472 could be atypical, daily faecal samples were collected from three other vaccinated infants; 4 days after vaccination all three subjects were excreting virus having C at position 472 (Table 2).

Table 2 Base at position 472 of the poliovirus genome in primary vaccinees

Day post-vaccination	KT1	Vaccinee KT2	KT3
1	U	*	U
2	U/C	*	U
3	C	*	C
4	C	C	C

Faecal samples were taken daily from three vaccinated infants less than 1 yr old, who received vaccine of the same origin as DM (Table 1).

* Isolates of poliovirus type 3 not available.

Virus with C at position 472 is thus selected for rapidly and consistently following administration of vaccine, the short period of time before the appearance of the first isolate with C probably representing only a few cycles of virus replication.

The basis of the selective advantage conferred by the presence of C at position 472 and the function of the 5' noncoding region are unclear. The degree of sequence homology between the three poliovirus serotypes in the 5' noncoding region is high (74%)¹⁶. Moreover, this is the region that is most conserved between the genomes of polioviruses and rhinovirus type 14 (ref. 17). This strong conservation suggests that the region has functional significance. One possibility is that some of the short open reading frames present in this region are functional and that their products remain undetected. Other possible functions of this 5' region include replicase binding on the negative-strand RNA, packaging of RNA into new virus capsids, uncoating and control of the initiation of translation. Such functions could conceivably depend on RNA secondary structure.

Figure 1 shows a prediction of the secondary structure of the RNA^{18,19} of the Sabin vaccine virus and of isolate P3/119; the change from U to C at position 472 in P3/119 causes a substantial rearrangement, so that two additional stem-and-loop structures are formed. A similar structure to that shown for P3/119 was also predicted for the virulent vaccine progenitor P3/Leon/37.

We have shown that for poliovirus type 3, a single nucleotide change of U to C at position 472 in the noncoding region of the viral genome provides a selective advantage for growth in the human gut. This change is associated with an increase in neurovirulence of the virus although it does not, on its own, give rise to a fully neurovirulent phenotype. Neither the nature of the selective environment nor the basis of the selective advantage conferred by the change is known.

It is of interest that, according to the particular RNA folding algorithm used in the present study, the U to C mutation appears to have a profound effect on the predicted secondary structure

of the virus RNA, but it remains to be established whether the secondary structures presented and the differences between them have any biological significance. It is possible that analogous changes are important in the attenuation of poliovirus types 1. The type 1 vaccine strain, P1/LSc, shows seven differences compared with its neurovirulent parent, P1/Mahoney, in the noncoding region of its genome¹⁶, and consequently some major rearrangements in the secondary structure of the RNA were predicted by this particular algorithm, including some in the region of position 472.

We thank Dr Clayton Naeve, St Jude's Children's Hospital, Memphis, Tennessee, for the synthesis of several oligonucleotide primers used in this study. We also thank Dr Frank Taffs, Dr Christine Miller, Susan Marsden, Anne John, Susan Fraser and David Minor for their contributions.

Received 22 November 1984; accepted 28 January 1985.

1. Assaad, F. & Cockburn, W. C. *Bull. Wild Hlth Org.* **60**, 231-242 (1982).
2. Minor, P. D. *J. Virol.* **34**, 73-84 (1980).
3. Minor, P. D. *J. gen. Virol.* **59**, 307-317 (1982).
4. Cann, A. J. *et al. Nucleic Acids Res.* **12**, 7787-7792 (1984).
5. Cossart, Y. E. *Br. med. J.* **1**, 162-163 (1977).
6. Kew, O. M., Nottay, B. K., Hatch, M. H., Nakano, J. H. & Obijeski, J. F. *J. gen. Virol.* **56**, 337-347 (1981).
7. Stanway, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 1539-1543 (1984).
8. Stanway, G. *et al. Eur. J. Biochem.* **135**, 529-533 (1983).
9. WHO Collaborative Study. *J. biol. Standard.* **9**, 163-184 (1981).
10. Evans, D. M. A., Minor, P. D., Schild, G. C. & Almond, J. W. *Nature* **304**, 459-462 (1983).
11. Toyoda, H. *et al. J. molec. Biol.* **174**, 561-585 (1984).
12. Magrath, D. L., Boulger, L. R. & Hartley, E. G. *10th Eur. Symp. Poliomyelitis*, 334-345 (EAP, 1964).
13. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
14. *Tech. Rep. Ser. Wild Hlth Org.* **687**, 130 (1983).
15. *Tech. Rep. Ser. Wild Hlth Org.* **687**, 134 (1983).
16. Nomoto, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 5793-5797 (1982).
17. Stanway, G., Hughes, P. J., Mountford, R. C., Minor, P. D. & Almond, J. W. *Nucleic Acids Res.* **12**, 7859-7875 (1984).
18. Jacobson, A. B., Good, L., Simonetti, J. & Zuker, M. *Nucleic Acids Res.* **12**, 45-52 (1984).
19. Shapiro, B. A., Maizel, J. V., Lipkin, L. E., Currey, K. & Whitney, C. *Nucleic Acids Res.* **12**, 75-88 (1984).

Suppression of leukaemia virus pathogenicity by polyoma virus enhancers

Brian Davis*, Elwood Linney† & Hung Fan*

* Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92717, USA

† Department of Microbiology, Duke University School of Medicine, Durham, North Carolina 27706, USA

The long terminal repeats (LTRs) of retroviruses contain sequences necessary for the initiation and termination of retroviral transcription. These sequences include promoter elements, transcriptional termination signals and transcriptional enhancer elements. The enhancer elements of Moloney murine leukaemia virus (M-MuLV) are localized in a tandemly repeated region (~75 base pairs (bp) long), which lies 5' to the CAT and TATA promoter elements in the U3 region of the LTR (see Fig. 1)¹. We have shown that the tandem repeats are required both for LTR promoter activity, as measured by transient expression assays, and for biological activity, as measured by production of infectious virus². Furthermore, they can be replaced by transcriptional enhancers from the F101 host-range mutant of polyoma virus without loss of function. We report here that the addition of the polyoma (PyF101) enhancers to the M-MuLV LTRs (either with or without the M-MuLV tandem repeats) results in complete loss of viral leukaemogenicity, even though the virus can replicate to high titres in tissue culture fibroblasts and can establish infection in animals.

Figure 1A shows the modified M-MuLV LTRs used in the present study. Mo+PyF101 was constructed by inserting the PyF101 enhancer element into the wild-type M-MuLV LTR at

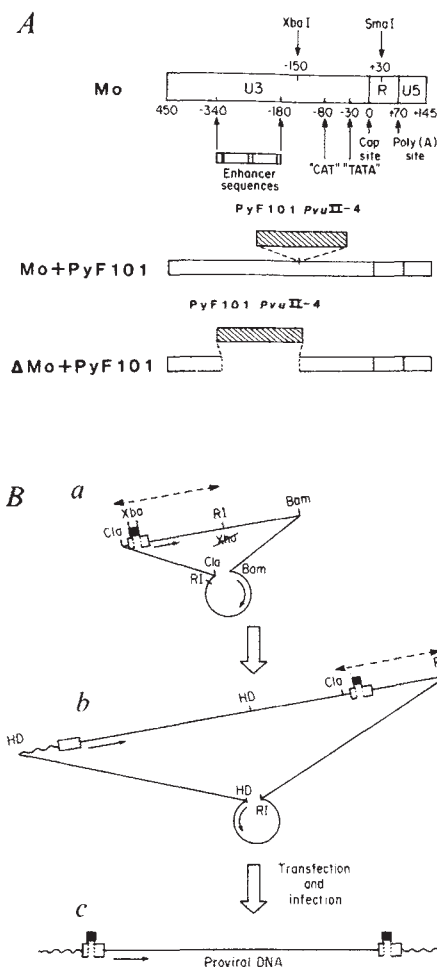


Fig. 1 A, M-MuLV LTRs used for tests of infectivity and pathogenicity. Mo, The wild-type M-MuLV LTR from a clone of circularly permuted proviral DNA, pMLV-1a^{1,24}; the sequences important in proviral transcription are indicated. The tandem repeats, implicated as enhancer elements^{14,25}, are located upstream (-341 to -182) of the promoter region. Mo+PyF101, LTR containing the PvuII-4 fragment of mutant polyoma virus PyF101 (ref. 26) inserted into the M-MuLV LTR at the XbaI site at position -150, 30 bp 3' of the M-MuLV tandem repeats. Mo+PyF101 was constructed by first adding XbaI linkers to the 199-bp PvuII-4 fragment², then inserting it into M-MuLV at the XbaI site. ΔMo+PyF101, LTR containing the PyF101 PvuII-4 fragment inserted into an M-MuLV LTR in which the tandem repeats (-335 to -150) have been deleted (see ref. 2 for details of construction). The orientation of the PyF101 PvuII-4 fragment relative to the Moloney promoter is identical in ΔMo+PyF101 and Mo+PyF101. B, The strategy used to construct M-MuLVs with modifications in the U3 region of the LTR. The ClaI to EcoRI fragment of ΔpMLV-C/R/B+PyF101 (a) was used to construct a complete M-MuLV recombinant DNA provirus containing a 3' ΔMo+PyF101 LTR (b) (see ref. 2 for details). The construct containing the Mo+PyF101 LTR was produced in a similar fashion. As the LTR alterations were all located within the U3 region, transfection of these constructs and subsequent infection of neighbouring cells would result in proviruses containing modified LTRs at both their 5' and 3' ends (c).

-150 bp, between the M-MuLV tandem repeats and the M-MuLV promoter. ΔMo+PyF101 comprises an insertion of the PyF101 enhancer into an M-MuLV LTR from which the M-MuLV enhancer elements (from -335 to -150 bp with respect to the RNA cap site) have been deleted². The biological activity of the LTRs was assessed by constructing a plasmid containing the complete M-MuLV proviral DNA sequence, with the altered LTR replacing the 3' LTR, transfecting this plasmid into NIH 3T3 cells and assaying for infectious virus² (see Fig. 1B for a

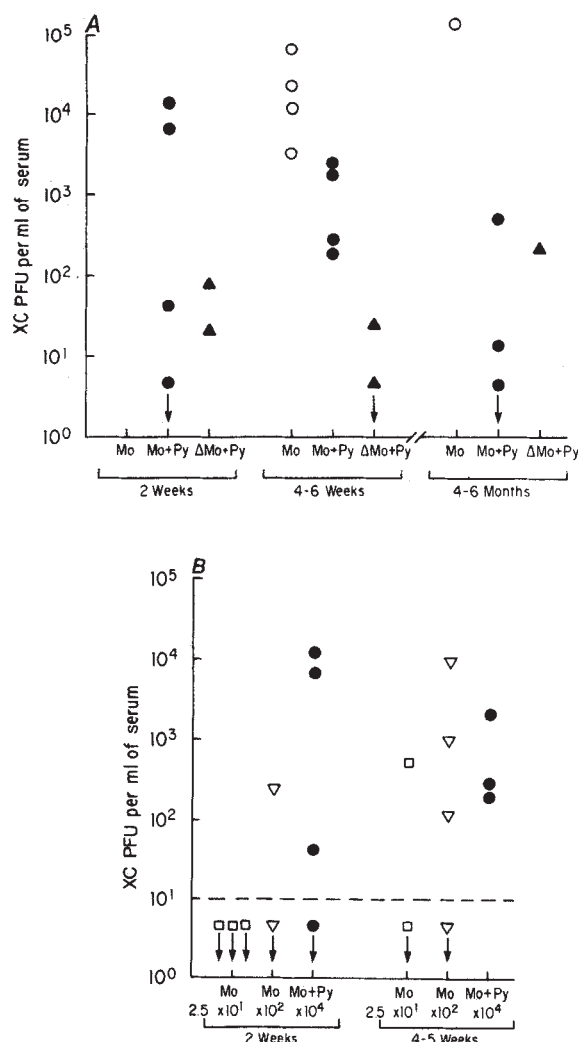


Fig. 2 Levels of circulating infectious virus in inoculated animals. Serum samples from mice injected i.p. 1-5 days after birth with either wild-type or polyoma-containing M-MuLVs were assayed by infection and titration on NIH 3T3 cells using the UV-XC assay²². Each data point in A and B shows the titration results obtained for a different sacrificed animal. Data points with downward-pointing arrows represent animals having virus levels undetectable in this assay ($<10^1$ XC PFU per ml of serum). A, Levels of virus in animals inoculated with 2.5×10^4 XC PFU of wild-type (○), Mo+PyF101 (●) or Δ Mo+PyF101 (▲) M-MuLVs. Titration results are given for animals at 2 weeks, 4-6 weeks, and 4-6 months after injection. B, Levels of virus in animals inoculated with 2.5×10^1 XC PFU of wild-type M-MuLV (□), 2.5×10^2 XC PFU of wild-type M-MuLV (▽), or 2.5×10^4 XC PFU of Mo+PyF101 M-MuLV (●). Results are shown for animals 2 weeks and 4-5 weeks post-injection.

brief outline of these procedures). Infectious virus containing either of the altered LTRs could be recovered (designated Mo+PyF101 M-MuLV and Δ Mo+PyF101 M-MuLV). Note that the alterations in these viruses lie in regions that do not encode protein, thus the virus particles are identical to wild-type M-MuLV.

To assess the relative infectivities of the Δ Mo+PyF101 and Mo+PyF101 viruses, virus produced by the transfected cultures was infected onto NIH 3T3 cells, which were then transferred until confluent infected. The amount of infectious virus produced by the cultures was assayed by titration of tissue culture supernatants on NIH 3T3 cells (Table 1). Cells infected with the two PyF101-containing viruses produced high amounts of infectious virus, although less than the amount produced by wild-type M-MuLV. As quantified by the reverse transcriptase

Table 1 Biological activity of M-MuLV containing altered LTRs

LTR	Titre of infectious virus	No. of virus particles (reverse transcriptase assay)	Infectivity-to-particle ratio
Mo	1.1×10^7	4.4×10^6	2.5
Mo+PyF101	1.6×10^6	3.0×10^6	0.53
Δ Mo+PyF101	5.0×10^5	3.1×10^6	0.16

Virus released from NIH 3T3 cells transfected with the M-MuLV proviral constructs containing altered LTRs was infected onto fresh NIH 3T3 cells. The freshly infected cells were transferred four times, after which all cells were infected; 24-h tissue culture supernatants were collected from 10^6 cells and assayed for concentration of infectious virus and for the number of virus particles present. The titre of infectious virus was quantified by titration of supernatants onto NIH 3T3 cells using the rat XC cell syncytial assay²²; the values shown represent XC PFU ml⁻¹. The number of virus particles present was determined by the reverse transcriptase assay, using poly(rA)/oligo(dT) as exogenously added template/primer²³; the values shown represent c.p.m. dTTP incorporated in a 75-min reaction for virus collected from 1 ml of tissue culture supernatant. Typical values for dTTP incorporation ranged from 4.0×10^5 to 6.5×10^5 c.p.m.; supernatant from uninfected NIH 3T3 cells yielded approximately 1% as much dTTP incorporation. Infectivity-to-particle ratios were calculated by dividing the infectious virus titre by the reverse transcriptase activity for each of the infected cell tissue culture supernatants.

assay (Table 1), cells infected with Δ Mo+PyF101 and Mo+PyF101 M-MuLVs produced almost the same number of virus particles as wild-type M-MuLV. To compare directly the infectivity of the polyoma-containing MuLVs with that of wild-type M-MuLV, infectivity-to-particle ratios were calculated by dividing the infectious virus titre by the reverse transcriptase activity present in each of the infected tissue culture supernatants (Table 1). Compared with wild-type M-MuLV, Mo+PyF101 virus showed a slight (fivefold) decrease in specific infectivity and Δ Mo+PyF101 virus a somewhat greater (15-30-fold) decrease. Thus, the polyoma-containing M-MuLVs are able to productively infect NIH 3T3 fibroblasts, and show only slightly decreased infectivity compared with wild-type M-MuLV.

The pathogenicity of wild-type, Δ Mo+PyF101 and Mo+PyF101 M-MuLVs was assessed by inoculating neonatal (1-3-day-old) NIH Swiss mice intraperitoneally (i.p.) with equal infectious units ($\sim 2.5 \times 10^4$ XC plaque-forming units, PFU) of each virus, then monitoring the mice for appearance of disease; any moribund animals were autopsied. Injection of six neonates with wild-type M-MuLV resulted in development of typical M-MuLV thymic leukaemia and death in all of the animals. Diseased animals showed enlarged thymuses, spleens or lymph nodes. The mean time until the disease became apparent was 3 months, with all mice developing the disease before 4 months. These results are in agreement with those obtained by us and others for M-MuLV³. In contrast, there was no evidence of disease in 10 neonates (from two litters) injected with Mo+PyF101 M-MuLV and in 17 neonates (from two litters) injected with Δ Mo+PyF101 M-MuLV, even 12-13 months after injection. Also, autopsies of animals inoculated with either virus showed no signs of abnormality or malignancy. These results suggest that Mo+PyF101 and Δ Mo+PyF101 M-MuLVs are non-leukaemogenic, even when inoculated at high titre.

One possible explanation for the lack of pathogenicity of the altered viruses is that they are unable to establish infection *in vivo*, even though they can replicate in tissue culture. To examine this possibility, circulating levels of infectious virus in inoculated animals were measured by titration of the blood serum on NIH 3T3 cells (Fig. 2A). Infectious XC-positive MuLV was recovered from the great majority of mice (13 out of 16) inoculated with Mo+PyF101 M-MuLV or Δ Mo+PyF101 M-MuLV, at time points between 2 weeks and 6 months after injection. XC-positive virus (although at low titre) was recovered even 5-6

months after injection with the polyoma-containing M-MuLVs, which further emphasizes that these viruses were able to establish and sustain multiple rounds of infection in the inoculated neonates. The low titres of virus present at later time points in Mo + PyF101 M-MuLV-injected animals (compared with animals inoculated with wild-type M-MuLV) may have resulted from an immune response in the absence of any malignant growth of infected thymoma cells.

Even though the polyoma-containing viruses were clearly able to establish infection *in vivo*, it was conceivable that the lack of leukaemogenicity might have resulted from a quantitative defect in the level of infection established. Multiple rounds of infection in the animal are probably necessary for leukaemogenesis by M-MuLV, shown by the eventual appearance of 'mink cell focus-inducing' recombinant MuLVs^{3,4}, and by the site specificity of proviral integration in tumour cells⁵⁻⁸. To investigate this possible defect in more detail, neonatal mice were inoculated with lower levels of wild-type M-MuLV which were still sufficient to induce leukaemia (2.5×10^1 and 2.5×10^2 XC PFU per animal) and the circulating levels of M-MuLV were compared with those in animals inoculated with concentrated Mo + PyF101 M-MuLV (Fig. 2B). At 2 weeks post-injection, the levels of viraemia in animals inoculated with the diluted stocks of wild-type M-MuLV were significantly lower (undetectable in several animals) than in animals inoculated with concentrated Mo + PyF101 M-MuLV. Even at 4-5 weeks after injection, the increased levels of viraemia in animals inoculated with wild-type M-MuLV were only equivalent to those in animals inoculated with Mo + PyF101 M-MuLV. Thus, the level of infection in the Mo + PyF101 M-MuLV-inoculated animals would have been more than sufficient to cause disease if the virus had been wild-type. Animals inoculated with Δ Mo + PyF101 M-MuLV also showed equivalent or higher levels of viraemia at 2 weeks post-inoculation than the animals inoculated with the diluted M-MuLV stocks.

The virus recovered from a 5-month animal inoculated with Mo + PyF101 M-MuLV was further characterized, since it was possible that recombination within the animal might have given rise to virus different from the input Mo + PyF101 M-MuLV. RNA and high relative molecular mass DNA were extracted from NIH 3T3 cells uniformly infected with the recovered virus and analysed by RNA dot-blot and DNA Southern blot analysis, respectively (Fig. 3A, B). RNA from the infected cells hybridized with a nick-translated PyF101 probe, confirming that M-MuLV containing the Mo + PyF101 LTR was still present in the animal 5 months after injection (Fig. 3A). DNA from cells infected with the recovered virus was digested with *Pvu*II and Southern blot hybridization was performed using an M-MuLV probe which detects a 1.88-kilobase (kb) fragment extending from -0.45 kb (relative to the cap site) in the Mo + PyF101 LTR to a *Pvu*II site at 1.43 kb on the proviral map (Fig. 3C). This *Pvu*II fragment is 199 bp larger in the Mo + PyF101 provirus than the equivalent fragment from wild-type M-MuLV provirus (1.68 kb) as a result of the insertion of the PyF101 sequences. Figure 3B shows that the *Pvu*II fragment in cells infected with the recovered virus was the same size as that in cells infected with the original Mo + PyF101 M-MuLV. Furthermore, no *Pvu*II fragments of different size were observed, indicating that no new M-MuLVs with recombinant LTRs arose in the animals inoculated with Mo + PyF101 M-MuLV.

These results suggest that the presence of polyoma virus enhancer sequences in the M-MuLV LTR abolishes the leukaemogenicity of the virus, even though it is still able to replicate in tissue culture fibroblasts and in the animal. The loss of pathogenicity did not result from an inability to establish infection in the animal, suggesting rather that it may reflect a change in tissue tropism of the virus resulting from addition of the polyoma enhancer sequences. The loss of leukaemogenicity may be a result of the polyoma-containing viruses being less able to infect the target cell(s) involved in development of M-MuLV lymphoma. There is accumulating evidence that certain transcriptional regulatory elements, including enhancers,

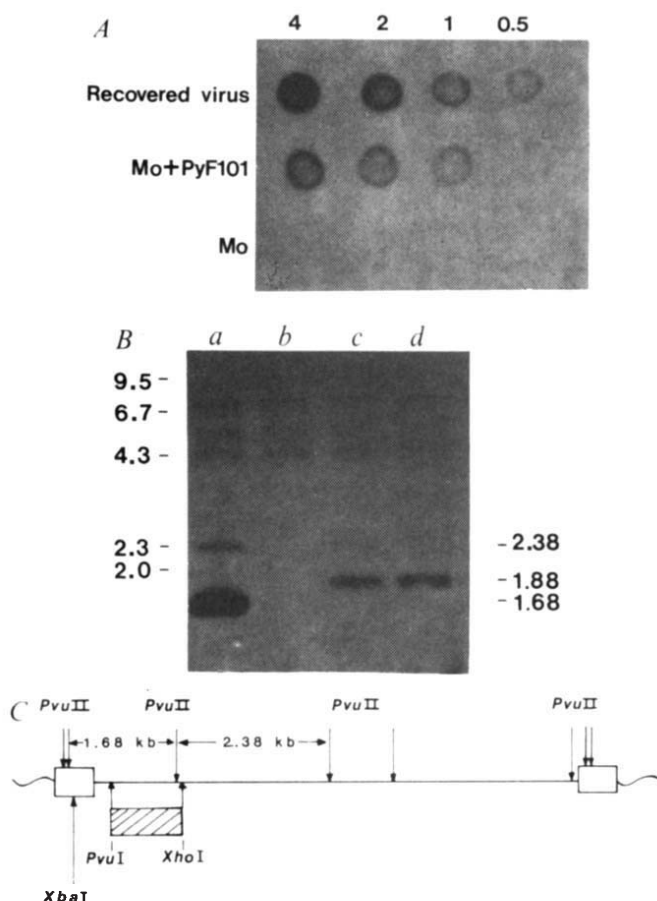


Fig. 3 Characterization of virus recovered from an animal inoculated with Mo + PyF101 M-MuLV. **A**, RNA dot-blot hybridization²⁷ of total cytoplasmic RNA²⁸ isolated from NIH 3T3 cells confluent infected with wild-type M-MuLV (Mo), Mo + PyF101 M-MuLV, or virus recovered from the bloodstream of a mouse 5 months after inoculation with Mo + PyF101 M-MuLV (recovered virus). The amounts of RNA spotted onto the nitrocellulose sheet²⁷ were as shown (4, 2, 1 and 0.5 μ g). The sheet was hybridized with ³²P-labelled nick-translated PyF101 *Pvu*II-4 fragment, and subjected to autoradiography. The probe hybridized with RNA from cells infected with Mo + PyF101 and with recovered virus, but not with RNA from cells infected with wild-type M-MuLV. RNA from all three cell lines hybridized with a labelled nick-translated M-MuLV probe (data not shown). **B**, Southern blot hybridization of *Pvu*II-digested genomic DNA of high relative molecular mass, isolated from NIH 3T3 cell lines infected with wild-type M-MuLV (**a**), Mo + PyF101 M-MuLV (**d**) or recovered virus (**c**), and from uninfected NIH 3T3 cells (**b**). The Moloney cell line used in this experiment, M-MuLV clone A9, has been described previously²³. Isolation of genomic DNA and blot-transfer hybridization were performed as described elsewhere^{29,30}. The hybridization probe was a 1.1-kb *Pvu*I to *Xho*I fragment from 0.42 to 1.55 kb, relative to the cap site, on the proviral map (**C**). Hybridization with *Pvu*II-digested M-MuLV clone A9 DNA revealed two fragments of 1.68 and 2.38 kb (**a**), which were not present in *Pvu*II-digested uninfected NIH 3T3 DNA (**b**). The weakly hybridizing fragments (~4-9 kb) present in both uninfected and infected NIH 3T3 cells represented endogenous M-MuLV proviral sequences present in uninfected mouse cells. Insertion of the 199-bp PyF101 *Pvu*II-4 fragment in the Mo + PyF101 LTR increased the size of the 1.68-kb M-MuLV fragment to 1.88 kb, as predicted (**d**). The 2.38-kb M-MuLV fragment, although difficult to identify in **d**, was recognizable in the original autoradiogram. The weak hybridization to the 2.38-kb fragment was expected, as only a small fraction of the nick-translated M-MuLV *Pvu*I to *Xho*I restriction fragment could hybridize with the 2.38-kb *Pvu*II fragment (see **C**). The hybridization pattern obtained for cells infected with the recovered virus (**c**) was indistinguishable from that of Mo + PyF101 M-MuLV-infected cells (**d**). **C**, The *Pvu*II restriction map of an integrated wild-type M-MuLV provirus, and the relative position of the nick-translated *Pvu*I to *Xho*I probe (hatched bar). The *Xba*I site in the LTR is also indicated.

can confer tissue specificity of expression^{2,9-14}; in fact, recent reports suggest that enhancer sequences in the LTR determine the thymotropism and pathogenicity of murine leukaemia viruses¹⁵⁻²¹.

We have found that loss of leukaemogenicity occurs not only for an M-MuLV LTR in which the M-MuLV tandem repeats have been replaced by the PyF101 enhancer ($\Delta\text{Mo} + \text{PyF101}$), but also for an M-MuLV LTR in which PyF101 enhancers have been added to a wild-type LTR still containing the M-MuLV enhancers ($\text{Mo} + \text{PyF101}$). The behaviour of $\text{Mo} + \text{PyF101}$ M-MuLV suggests that the PyF101 enhancer sequences exerted a dominant effect over the M-MuLV enhancer sequences present in the $\text{Mo} + \text{PyF101}$ LTR.

This work was supported by NIH grants R01 CA32454 and CA32455 to H.F. and an ACS Postdoctoral Fellowship to B.D.

Received 14 January; accepted 4 February 1985.

1. Van Beveren, C., Rands, E., Chattopadhyay, S. K., Lowy, D. R. & Verma, I. M. *J. Virol.* **41**, 542-556 (1982).
2. Linney, E., Davis, B., Overhauser, J., Chao, E. & Fan, H. *Nature* **308**, 470-472 (1984).
3. Vogt, M. *Virology* **93**, 226-236 (1979).
4. Hartley, J. W., Wolford, N. K., Old, L. J. & Rowe, W. P. *Proc. natn. Acad. Sci. U.S.A.* **74**, 789-792 (1977).
5. Li, Y., Holland, C. A., Hartley, J. W. & Hopkins, N. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6808-6811 (1984).
6. Corcoran, L. M., Adams, J. M., Dunn, A. R. & Cory, S. *Cell* **37**, 113-122 (1984).
7. Steffen, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2097-2101 (1984).
8. Cuypers, H. T. *et al. Cell* **37**, 141-150 (1984).
9. Banerjee, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
10. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
11. Mercola, M., Wang, X.-F., Olsen, J. & Calamne, K. *Science* **221**, 663-665 (1983).
12. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. *Nature* **306**, 557-561 (1983).
13. Chen, I. S. Y., McLaughlin, J. & Golde, D. W. *Nature* **309**, 276-279 (1984).
14. Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6453-6457 (1982).
15. DesGroselliers, L., Rassart, E. & Jolicoeur, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4203-4207 (1983).
16. Chatis, P. A. *et al. J. Virol.* **52**, 248-254 (1984).
17. Lenz, J. *et al. Nature* **308**, 467-470 (1984).
18. Celander, D. & Haseltine, W. A. *Nature* **312**, 159-162 (1984).
19. DesGroselliers, L. & Jolicoeur, P. *J. Virol.* **52**, 448-456 (1984).
20. DesGroselliers, L. & Jolicoeur, P. *J. Virol.* **52**, 945-952 (1984).
21. Oliff, A., Signorelli, K. & Collins, L. *J. Virol.* **51**, 788-794 (1984).
22. Rowe, W. P., Pugh, W. E. & Hartley, J. W. *Virology* **42**, 1136-1139 (1970).
23. Fan, H., Jaenisch, R. & MacIsaac, P. *J. Virol.* **28**, 802-809 (1978).
24. Berns, A. J. M. *et al. J. Virol.* **36**, 254-263 (1980).
25. Jolly, D. J., Esty, A. C., Subramani, S., Friedmann, T. & Verma, I. M. *Nucleic Acids Res.* **11**, 1855-1872 (1983).
26. Fujimura, F. K., Deininger, P. L., Friedmann, T. & Linney, E. *Cell* **23**, 809-814 (1981).
27. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
28. Fan, H. & Baltimore, D. *J. molec. Biol.* **80**, 93-117 (1973).
29. Bachelier, L. T. & Fan, H. *J. Virol.* **30**, 657-667 (1979).
30. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).

Expression of enhanced levels of small RNA polymerase III transcripts encoded by the B2 repeats in simian virus 40-transformed mouse cells

Karambir Singh, Michael Carey, Sentob Saragosti & Michael Botchan

Department of Molecular Biology, University of California, Berkeley, California 94720, USA

Although specific viral genes are known which are sufficient to transform certain types of cells, viral transformation has been shown to be accompanied by a change in the level of expression of cellular genes. Assuming that this latter class of genes is involved in some way in the establishment of the transformed state, we and others^{1,2} have sought to characterize them. To this end, we constructed a complementary DNA library from a simian virus 40 (SV40)-transformed mouse cell line and screened this library for induced genes. Here we focus on the characterization of a specific cDNA clone called 10A5. We show that the clone 10A5 contains a B2 repeat^{3,4}, and that the levels of the small heterogeneously sized B2 cytoplasmic RNAs homologous to 10A5 are enhanced in transformed cells. Our studies show that these RNAs are a specific

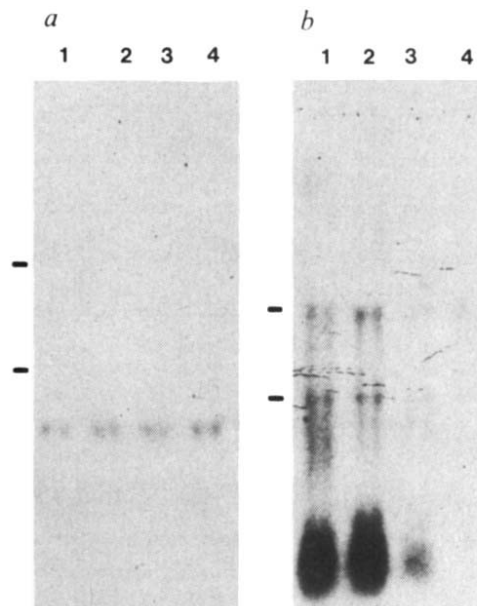


Fig. 1 Northern blot analysis of transcripts with homology to 8A2, a non-differentially expressed cDNA clone that serves as a control (a), and 10A5 (b). The lanes contain RNA from cells grown to different densities. Each lane contains 20 μg of cytoplasmic RNA. The lanes are: 1, SVLTR-1²², cell density per 90 cm plate = 2.3×10^6 ; 2, SVLTR-1, cell density 18×10^6 ; 3, NIH 3T3, cell density, 2.7×10^6 ; 4, NIH 3T3, cell density, 5.5×10^6 . Lanes 1 and 3 represent non-confluent cells and lanes 2 and 4 represent confluent cells. The positions of the 28S and 18S rRNAs are represented by lines on the left of each panel. In overexposed autoradiograms of total cytoplasmic RNA, some hybridization to both 18S and 28S RNA is detectable. The levels of the small RNAs homologous to 10A5 shown in b are about fivefold greater in the transformed cell than in the non-confluent parental cell and about 20-fold greater in the transformed cell than in the confluent parental cell.

Methods. To isolate cDNA clones representing differentially expressed genes such as 10A5, a plus/minus screen was performed on $\approx 10,000$ phage from the SVLTR-1 $\lambda\text{gt}10$ cDNA library. The cDNA probes used in the screening represented the poly(A)⁺ RNA of either the NIH 3T3 or SVLTR-1 cells. Those cDNAs that consistently revealed a differential expression were isolated, subcloned into pBR322 and used as probes of Northern blots. Cells were grown in Dulbecco's modified minimal essential medium (DMEM) containing 10% calf serum. For the isolation of cytoplasmic RNA, $\approx 10^8$ cells were washed with phosphate-buffered saline, and then lysed by resuspending the cells in 1 ml of 0.5 M NaCl, 10 mM Tris (pH 7.5), 1.5 mM MgCl_2 , 0.65% NP40. After incubation at 0 $^\circ\text{C}$ for 10 min, an equal volume of 7 M urea, 0.35 M NaCl, 10 mM Tris (pH 7.5), 10 mM EDTA, 1% SDS was added to the supernatant. This was extracted twice with phenol/chloroform/isoamyl alcohol (50:48:2) containing 0.1% 8-hydroxyquinoline, extracted twice with chloroform/isoamyl alcohol (24:1) and ethanol precipitated. For Northern blot analysis, 20 μg of cytoplasmic RNA was dissolved in sample buffer containing 1 $\times$ running buffer (RB, 20 mM MOPS (pH 7.0), 1 mM EDTA, 5 mM NaAc), 50% (v/v) formamide, and 2.2 M formaldehyde. The RNA was then heated for 5 min at 65 $^\circ\text{C}$ and electrophoresed in 1-1.5% agarose gels containing 1 $\times$ RB and 2.2 M formaldehyde. The position of the ribosomal RNA was visualized by acridine orange staining and the RNA was transferred to nitrocellulose paper in 20 $\times$ SSC. The filters were air dried and baked for 2 h *in vacuo* at 80 $^\circ\text{C}$. Hybridization conditions were as described elsewhere^{23,24}. Plasmid probes used were labelled by nick-translation to a specific activity of $1-2 \times 10^8$ c.p.m. per μg .

class of polymerase III transcripts containing a highly conserved 5' end. SV40 transformation results, therefore, in the activation of specific polymerase III transcripts as well as specific polymerase II transcripts.

Figure 1 is a Northern blot showing that the small RNAs homologous to 10A5 are more abundant in the SV40-transformed cell line than in the parental NIH 3T3 cell line. The

a

```

10      20      30      40      50      60      70      80      90      100
CCGCTAAATA AGAATGTAAG GGTGAGGCT TACTAGTAT GTTTTGTGT GTCAGTGA CAAGGCTCA ATTGTACTGT CTAGTTTGT GTCAACTTGA

110     120     130     140     150     160     170     180     190     200
CACAGCTGGA GTTATCAGAG AGAAGGACC TTCAGTTGAG GAATGCTCT CATGAGATCC AACTGTAAAG CATTTTCTCA ATTAGTGATC AAGGGGGAAA

210     220     230     240     250     260     270     280     290     300
GGCCCTTGT GGGTGGGACC ATCTCTGGGC TGGTATGCTT GGGTTTATA AGAGAGCAGG CTGAGCAGGC CAGGGGAGGC AAGCACTGAA AGAACATCCC

310     320     330     340     350     360     370     380     390     400
TCCATGAGCT CTGATCAGC TCTCTCTTTC TGACCTGCTT GAGTTCCAGT CCTGATCTCT TTGGTGACCA ACAGCAGTAT GGAATGTGAA GCGAATATA

410     420     430     440     450     460     470     480     490     500
CCCTTCTCTC CCCAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA GTGCTCTGG AAGGTCCAGA GTTCAATCC CAGCAACCC ATGGTGGCTC

510     520     530     540     550     560     570     580
ACAACCATCT GTAAACAGAT CTGACTCCCT CTCTGGAGT GTCTGAAGAC AGCTACAGTG TACTTACATA TAATTAATTA A]

```

b

```

B2      A-Box      B-Box
GGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCACTGAC TGCTCTTGGAGGCTGAGTTCAGTTCACCAACACATGGTGG
10λ5.1  TGCTC-TGGAGGCTGAGTTCAGTTCACCAACACATGGTGG

B2      CTACACACCATCGTAAAGATCTGATGCGCCCTCTCTGGAGTGTCTGAAGACAGCTACAGTCTACTTACATATAAATAA
10λ5.1  CTACACACCATCTGTAAACAGATCTGACTCCCTCTCTGGAGTGTCTGAAGACAGCTACAGTCTACTTACATATAAATAA

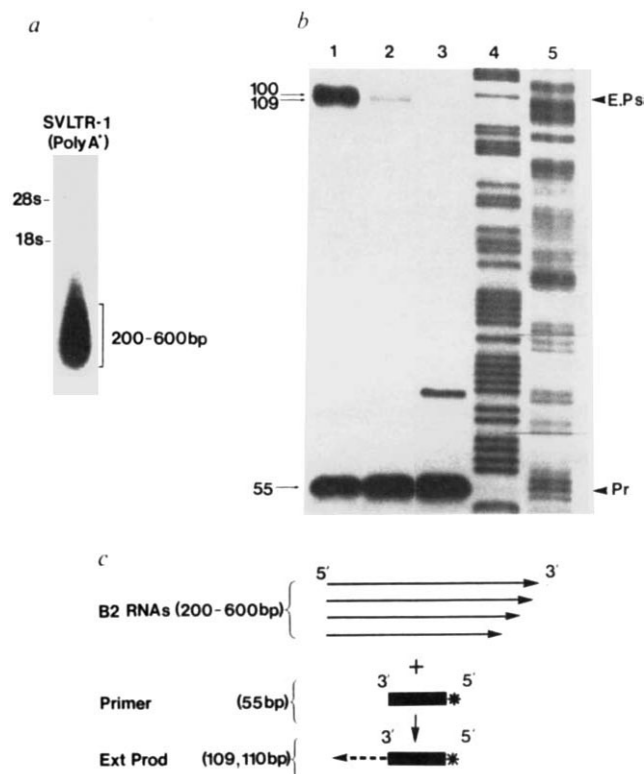
```

Fig. 3 **a**, Northern blot analysis of *B2* transcripts homologous to 10λ5 in 1 μg of poly(A)⁺, cytoplasmic SVLTR-1 RNA. **b**, Primer extension analysis of SVLTR-1 cytoplasmic RNA using a *B2* gene specific probe. 1 μg (lane 1) and 100 ng (lane 2) of RNA were subjected to primer extension as described in **c**. Lane 3 is a mock reaction with no RNA. The extension products were sized alongside a sequencing ladder of known sequence. The *Bam*HI/*Hind*III fragment of the plasmid pXbs-73x201 (a *Xenopus borealis* 5S clone) was labelled at the *Bam* site using polynucleotide kinase and [γ ³²P]ATP. The fragment was then subjected to Maxam-Gilbert sequencing reactions²⁵ for purines (lane 4) and pyrimidines (lane 5). The band which migrated just above the primer in lane 3 is double-stranded primer which remained undenatured. Pr denotes primer and E.P.s denotes extension products. **c**, A schematic illustration of the *B2* primer extension procedure. The 55-nucleotide *Bgl*II/*Ava*II coding strand fragment of 10λ5 (Fig. 2a, nucleotides 467-521), 5' labelled at the *Bgl*II site, was prepared by standard procedures. This fragment was hybridized to total SVLTR-1 cytoplasmic RNA. The hybridized primer was then extended to the 5' end of the RNA by reverse transcriptase and deoxy-nucleotide triphosphates (dNTPs) to generate two major extension products of 109 and 110 nucleotides. More than 80% of the extension products are associated with this doublet. The 3' heterogeneity of *B2* RNAs is illustrated by the different-sized arrows.

Methods. Primer extension reactions were performed as described by McKnight *et al.*²⁶ with minor modifications. RNA samples were mixed with 5 × 10⁴ c.p.m. (0.01-0.05 pmol) of single-stranded labelled DNA primer in 300 mM NaCl, 10 mM Tris pH 7.5 and 1 mM EDTA. The mixture was heated at 85 °C for 10 min, then annealed at 55 °C for 2.5 to 3.0 h. Following the hybridization, samples were incubated with 100 U ml⁻¹ of avian myeloblastosis virus reverse transcriptase in a buffer containing 60 mM NaCl, 10 mM Tris pH 7.5, 10 mM dithiothreitol (DTT), 1 mM dNTPs, 8 mM MgCl₂, 50 μg ml⁻¹ actinomycin D and 1,000 U ml⁻¹ RNasin. These reactions were incubated at 42 °C for 1 h and terminated by phenol extraction. The reactions were ethanol precipitated and resuspended in 90% formamide dye mix, heated at 90 °C for 3 min and loaded onto 8% polyacrylamide sequencing gels. Gels were autoradiographed for 12 h with a Cronex intensifying screen.

augmentation ranges over 5-20-fold depending on the growth state of the parental cells. These small RNAs are also detected by 10λ5 in Northern blots of poly(A)⁺-selected RNAs (Fig. 3a). We find that these small RNAs are augmented in three other SV40-transformed mouse NIH 3T3 cell lines and in bovine papilloma virus- and methylcholanthrene-transformed cell lines (data not shown). Southern⁵ hybridization analysis of genomic DNA from various mouse cell lines using 10λ5 as a probe always showed an intense smear pattern, indicating that 10λ5 contains a highly repetitive element.

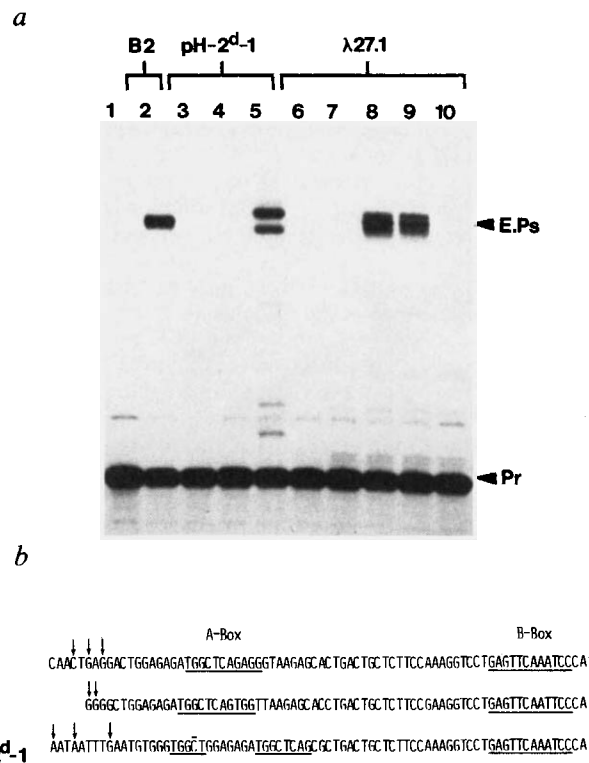
Fig. 2 **a**, The nucleotide sequence of 10λ5 determined by the chemical degradation procedure of Maxam and Gilbert²⁵. The *B2* repeat sequences are underlined. The upstream sequence shows no extensive homology to any known sequence. **b**, Comparison of the sequence of the *B2* repeat in 10λ5 with the rodent *B2* consensus sequence⁴. Asterisks represent differences between the 10λ5 repeat and the *B2* consensus sequence. The sequences representing the A and the B boxes of the split polymerase III promoter¹⁵ are underlined. The 10λ5 *B2* sequence lacks both the first 39 nucleotides at the 5' end of the repeat, including the putative A box of the split polymerase III promoter, and 15-20 bp from the heterogeneous, A-rich 3' end. However, over the remaining 131 bp, our sequence differs by only 6% from the consensus sequence.



The nucleotide sequence of 10λ5 is shown in Fig. 2a. The cDNA is 581 base pairs (bp) long, and at one end there is an incomplete copy of a member of the highly repetitive mouse *B2* family^{3,4}. Family members are about 190 bp in size and are highly conserved in all rodent species. *B2* RNAs were initially observed to be present at higher levels in both Ehrlich carcinoma and MOPC cells than in mouse liver cells⁶. The reports of Scott *et al.*¹ and ourselves have now added significance to this initial observation by independently showing that these small RNAs are induced in various transformed mouse cells.

Fig. 4 a, Primer extension analysis of RNA transcribed *in vitro* with partially purified HeLa cell transcription factors and cloned DNAs containing *B2* elements. The two clones studied were pH-2^d-1 and λ 27.1. The plasmid pH-2^d-1 contains a *B2* element in the 3'-untranslated region of the *H-2D* mouse class I gene of the major histocompatibility complex (MHC)⁸ and the λ clone λ 27.1 contains a mouse class I MHC pseudogene which has a *B2* element in the third intron⁷. Lane 1, a mock reaction with the primer alone; lane 2, the extension products for 500 ng of SVLTR-1 RNA; lanes 3–5, the extension products of RNAs generated by incubating pH-2^d-1 with polymerase III and TF-B (lane 3), polymerase III and TF-C (lane 4) and polymerase III, TF-B and TF-C (lane 5); lanes 6–10, the extension products of RNAs produced by incubating λ 27.1 with polymerase III and TF-B (lane 6), polymerase III and TF-C (lane 7), polymerase III, TF-B and TF-C (lane 8), polymerase III, TF-B, TF-C and 2 μ g ml⁻¹ of α -amanitin (lane 9) and polymerase III, TF-B, TF-C and 200 μ g ml⁻¹ of α -amanitin (lane 10). It is not clear why the *in vivo* pattern is more homogeneous than the *in vitro* patterns. Longer exposures of this autoradiogram show that the *in vivo* starts as seen in lane 2 contain other minor bands within 2–3 bp both 5' and 3' to the starts shown. Pr denotes primer and E.P.s denotes extension products. **b**, A sequence comparison of the *B2* elements transcribed *in vitro* with the *B2* consensus sequence. Both pH-2^d-1 and λ 27.1 are aligned with the *B2* consensus sequence by their B box homologies. The A and B box sequence homologies are underlined and the *in vitro* and *in vivo* initiation sites are denoted by arrows.

Methods. A HeLa cell S100 extract was prepared as described by Weil *et al.*²⁷. The transcription factor (TF) assays and the first step of the purification were modifications of those described by Segall *et al.*¹⁶. The extract was loaded onto a phosphocellulose column (P11 Whatman) pre-equilibrated in buffer A (20 mM HEPES pH 7.5, 20% glycerol, 0.2 mM EDTA and 0.5 mM DTT) and 100 mM KCl. TF-B activity was eluted with buffer A and a linear salt gradient of 100–350 mM KCl. Active fractions were pooled and dialysed for 8 h against buffer A and 50 mM KCl. The dialysate was loaded onto a DEAE-cellulose (DE-52 Whatman) column pre-equilibrated in the same buffer. The column was developed with buffer A and a linear salt gradient of 50–300 mM KCl. At this point, TF-B was 100-fold purified and contained <5% contaminating TF-C activity and no detectable TF-A activity. After the elution of TF-B, the phosphocellulose column was washed extensively with buffer A and 350 mM KCl. TF-C activity was then step eluted with buffer A and 600 mM KCl. Protein-containing fractions were pooled and dialysed for 8 h against buffer A and 200 mM KCl. The dialysate was loaded onto a native DNA cellulose column pre-equilibrated in the same buffer. The column was developed with buffer A and 1 M KCl. Protein-containing fractions were dialysed for 8 h against buffer A and 200 mM KCl. At this point, TF-C was purified 250-fold and contained <5% contaminating TF-B and no detectable TF-A activity. Both TF-B and TF-C contained endogenous RNA polymerase III as measured by a nonspecific template assay²⁸. The addition of purified RNA polymerase III had little effect on TF activity in the conditions used in our experiments. The standard transcription reactions were in a final vol. of 50 μ l and contained 20 mM HEPES pH 7.5, 10% glycerol, 5 mM MgCl₂, 66 mM KCl, 0.5 mM DTT, 200 μ M of rATP, CTP, GTP and UTP, 0.8 μ g of plasmid or λ clone DNA, 2.5 μ l TF-C and 12 μ l TF-B. Incubations were carried out at 23 °C for 1 h and were terminated by phenol extraction. The reactions were ethanol precipitated and the pellet was resuspended in 10 mM Tris, 1 mM EDTA. The RNA was then assayed by primer extension analysis as described in Fig. 3 legend. Primer extension products were sized by electrophoresis alongside a sequencing ladder of known sequence.



In the cytoplasm, *B2* sequences are found mainly in RNAs of low relative molecular mass that are not associated with polysomes⁶, but can also be present in larger, polysome-associated mRNAs at the 3'-untranslated end of the transcript⁷⁻¹². The cloned *B2* repeat in 10 λ 5 appears part of a larger transcript, as there is 450 bp of sequence 5' to the *B2* repeat. Figure 2b compares the 10 λ 5 *B2* sequence with the consensus *B2* sequence⁴.

The homologies to the polymerase III promoter sequences¹³ within the *B2* consensus sequence suggested that the small RNAs may be polymerase III transcripts. The heterogeneous nature of these RNAs could be a result of faithful initiation but irregular termination of transcription. Our examination of the sequence of a number of *B2* elements shows that only a subset contains an efficient polymerase III terminator—a cluster of four or more T residues surrounded by G+C-rich sequences¹⁴. Accordingly, one would then expect these cytoplasmic *B2* RNAs to have similar 5' ends and these ends to map near the 5' end of the element, because all polymerase III transcripts except for 5S RNA start 10–20 nucleotides upstream from the A box¹⁵.

We tested this hypothesis using primer extension. As a primer we used a 55-bp *B2*-specific fragment from 10 λ 5. The 5' end of the primer is located 31 bp downstream from the putative A box in the *B2* consensus sequence and 55 bp downstream from the 5' end of the element (see Fig. 2b). We annealed this primer to total cytoplasmic RNA and extended the hybrids with reverse transcriptase as shown schematically in Fig. 3c. We find two

major extension products 109 and 110 bp in size (Fig. 3b), which shows that the 5' ends of the mature cytoplasmic *B2* RNAs are located 12 and 13 nucleotides upstream from the A box. This result strongly suggests that the small cytoplasmic *B2* RNAs are polymerase III transcripts.

To corroborate this conclusion, we used a partially purified polymerase III transcription system to transcribe two clones, pH-2^d-1 (ref. 8) and λ 27.1 (ref. 7), which contained *B2* elements. Transcription of polymerase III genes with split promoters such as transfer RNA genes and the adenovirus virus-associated (VA) genes requires, in addition to polymerase III, transcription factors B and C (TF-B and TF-C)¹⁶. Primer extension analysis revealed that neither pH-2^d-1 or λ 27.1 was an effective template for polymerase III unless the reaction mix also contained both TF-B and TF-C (Fig. 4a, lanes 3–5 for pH-2^d-1 and lanes 6–8 for λ 27.1). The 5' ends of the *B2* RNAs produced *in vitro* (Fig. 4a, lanes 5, 8) align closely with the 5' ends of the *in vivo* *B2* RNAs (Fig. 4a, lane 2). Figure 4b summarizes these *in vitro* and *in vivo* start positions. Finally, λ 27.1 was incubated with polymerase III, TF-B and TF-C in the presence of 2 μ g ml⁻¹ (Fig. 4a, lane 9) and 200 μ g ml⁻¹ (Fig. 4a, lane 10) of α -amanitin. The lower level inhibits polymerase II transcription while the higher level inhibits polymerase III²⁹. Whereas 2 μ g ml⁻¹ of α -amanitin has no effect on *B2* transcription, 200 μ g ml⁻¹ is completely inhibitory.

The transcription factor requirements and the α -amanitin sensitivity indicate clearly that the mouse *B2* elements we have

studied are transcribed by polymerase III *in vitro*. In this respect, the mouse B2 family is equivalent to various other *Alu*-like sequences known to be polymerase III templates *in vitro* (for review see ref. 17). Our primer extension analysis of cytoplasmic RNA shows that these repeats are also functioning as expressed polymerase III genes *in vivo*. B2 RNA sequences can therefore arise as parts of polymerase II transcripts when the repeat is present within a polymerase II transcription unit, or they can be expressed independently by their own polymerase III promoters.

We have also studied 5S RNA gene expression and find no significant difference in the cytoplasmic levels of the 5S RNAs between the parental cell line and the SV40-transformed derivative (data not shown). Thus, there is no generalized induction of polymerase III transcription in SV40-transformed cells. Evidence for the regulation of specific polymerase III transcripts already exists for the expression of the oocyte-specific 5S genes in *Xenopus* (for review see ref. 18), and the brain-specific expression of the rat ID sequences^{19,20}. Moreover, it has recently been shown that E1A induction of adenovirus VA RNAs is at least in part due to elevated levels of polymerase III transcription factor C (W. K. Hoeffler and R. G. Roeder, and S. Yoshinaga and A. Berk, personal communications). If the induction of the B2 RNAs is at the level of transcription, this would suggest the presence of a polymerase III transcription factor(s) whose activity is sensitive to growth conditions and oncogenic transformation. It may be that the majority of the B2 repeats, like similar highly repeated *Alu*-like sequences, have no useful function in the cell and have simply keyed themselves into a growth-sensitive polymerase III transcription system. Thus, other templates for this system could also exist and these may have physiologically more important roles for the rapidly dividing cell. It will therefore be interesting to study how other polymerase III transcripts such as the small 7SL RNA involved in secretion²¹, tRNAs or 4.5S RNAs are affected by transformation.

We thank Drs P. Kourilsky, R. Goodenow and D. Bogenhagen for plasmid pH-2^d-1, phage λ 27.1 and plasmid pXbs-73x201, respectively, Drs D. Brutlag and F. Doolittle for conducting computer searches for DNA sequence homology to the clone 10 λ 5 and Dr N. Cozzarelli for help in monitoring and isolating polymerase III transcription factors and helpful comments on the manuscript. S.S. was supported by a fellowship from La Ligue Nationale Française contre le Cancer. This work was supported by grant CA 30490 from the National Cancer Institute and grant MV-91 from the American Cancer Society to M.B., and by NIH grant GM 31657 to N. Cozzarelli.

Received 18 October 1984; accepted 3 January 1985.

1. Scott, M. R. D., Westphal, K. & Rigby, P. W. J. *Cell* **34**, 557-567 (1983).
2. Schutzbank, T., Robinson, R., Oren, M. & Levine, A. J. *Cell* **30**, 481-490 (1982).
3. Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P. & Georgiev, G. P. *Nucleic Acids Res.* **6**, 697-713 (1979).
4. Krayner, A. S. *et al. Nucleic Acids Res.* **10**, 7461-7475 (1982).
5. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
6. Kramerov, D. A., Lekakh, I. V., Samarina, O. P. & Ryskov, A. P. *Nucleic Acids Res.* **10**, 7477-7491 (1982).
7. Steinmetz, M. *et al. Cell* **25**, 683-692 (1981).
8. Lalanne, J. L. *et al. Nucleic Acids Res.* **10**, 1039-1049 (1982).
9. Reyer, A. A., Schold, M. & Wallace, R. B. *Immunogenetics* **16**, 1-9 (1982).
10. Moore, K. W. *et al. Science* **215**, 679-682 (1982).
11. Brickell, P. M. *et al. Nature* **306**, 756-760 (1983).
12. Ryskov, A. P., Ivanov, P. L., Kramerov, D. A. & Georgiev, G. P. *Nucleic Acids Res.* **11**, 6541-6558 (1983).
13. Traboni, C., Gilierto, G. & Cortese, R. *EMBO J.* **1**, 415-420 (1982).
14. Bogenhagen, D. F. & Brown, D. D. *Cell* **24**, 261-270 (1981).
15. Ciliberto, G. *et al. Cell* **32**, 725-733 (1983).
16. Segall, J., Matsui, T. & Roeder, R. G. *J. biol. Chem.* **255**, 11986-11991 (1980).
17. Jelinek, W. R. & Schmid, C. W. A. *Rev. Biochem.* **51**, 813-844 (1982).
18. Korn, L. J. *Nature* **295**, 101-105 (1982).
19. Sutcliffe, J., Milner, R. J., Bloom, F. E. & Lerner, R. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4942-4946 (1982).
20. Sutcliffe, J., Milner, R. J., Gottesfeld, J. M. & Lerner, R. A. *Nature* **308**, 237-241 (1984).
21. Walter, P. & Blobel, G. *Nature* **299**, 691-698 (1982).
22. Krieger, M. & Botchan, M. *Molec. cell Biol.* **3**, 325-339 (1983).
23. Wahl, G. M., Stern, M. & Stark, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683-3687 (1979).
24. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1982).
25. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
26. McKnight, S. L., Davis, E. R., Kingsbury, R. & Axel, R. *Cell* **25**, 385-398 (1981).
27. Weil, P. A. *et al. J. biol. Chem.* **254**, 6163-6173 (1979).
28. Roeder, R. G. *J. biol. Chem.* **249**, 241-248 (1974).
29. Weinmann, R. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1790-1794 (1974).

ATP-dependent specific binding of Tn3 transposase to Tn3 inverted repeats

W. L. Wishart*, J. R. Broach* & E. Ohtsubo†

* Department of Molecular Biology, Princeton University, Princeton, New Jersey 08540, USA

† Department of Applied Microbiology, University of Tokyo, Bunkyo-ku, Tokyo, Japan 113

Transposons are discrete segments of DNA which are capable of moving from one site in a genome to many different sites^{1,2}. Tn3 is a prokaryotic transposon which is 4,957 base pairs (bp) long and encodes a transposase protein which is essential for transposition³⁻⁷. We report here a simple method for purifying Tn3 transposase and demonstrate that the transposase protein binds specifically to the ends of the Tn3 transposon in an ATP-dependent manner. The transposase protein binds to linear double-stranded DNA both nonspecifically and specifically; the nonspecific DNA binding activity is sensitive to challenge with heparin. Site-specific DNA binding to the ends (inverted repeats) of Tn3 is observed only when binding is performed in the presence of ATP; this ATP-dependent site-specific DNA binding activity is resistant to heparin challenge. Our results indicate that ATP qualitatively alters the DNA binding activity of the transposase protein so that the protein is able to bind specifically to the ends of the Tn3 transposon.

DNA sequence analysis reveals that the Tn3 transposase coding frame is 3,012 bp in length, which would allow for a protein of relative molecular mass (M_r) $\sim$ 113,000 (113K). This size estimate agrees well with results obtained in experiments where Tn3-specific proteins have been identified by radioactive labelling in the minicell strain DS410 (refs 7-9). The purification of Tn3 transposase was first reported by Fennwald *et al.*¹⁰. This protein exhibited a heat-sensitive nonspecific single- and double-stranded DNA binding activity. However, no additional enzymatic activities associated with the Tn3 transposase protein have been reported. To learn more about the Tn3 transposase, we devised a simpler purification scheme than that reported previously¹⁰. Our purification was facilitated by the observation that the wild-type transposase protein accumulates to $\sim$ 1.5% of total cell protein when a plasmid bearing a Tn3 repressor mutant (pMB8:Tn3^o)⁶, which transposes at a high frequency^{6,11}, is grown in minicell strain DS410^{12,13} (Fig. 1).

Briefly, our method consists of four essential steps: membrane extraction, polyethyleneimine precipitation, step dialysis and separation by phosphocellulose chromatography. The method is simple, rapid, gentle and yields milligram quantities of transposase protein (see Fig. 2). The identity of the transposase protein was confirmed by amino-acid analysis and N-terminal sequence analysis performed at the Brookhaven protein sequencing facility at Brookhaven National Laboratories (unpublished data), and our results agree with those reported by Ditto *et al.*¹⁴.

The transposase protein isolated in this manner binds non-specifically to linear DNA fragments (Fig. 3); this DNA binding activity can be abolished by the addition of heparin, a mucopolysaccharide which competes with DNA for DNA binding sites. We used an assay essentially identical to that described by Wishart *et al.*¹⁵. The transposase protein is added together with DNA fragments in 250 μ l of binding buffer (10 mM MgCl₂, 10 μ g ml⁻¹ bovine serum albumin, 100 mM KCl, 10 mM Tris-HCl pH 7.6, 0.1 mM dithiothreitol and 0.1 mM EDTA), ATP and heparin as described in Fig. 3 legend. The mixture is then incubated at 27 °C for 5 min and filtered through nitrocellulose filters. In our binding conditions, the transposase protein is retained quantitatively on the filters together with any DNA fragment to which the protein has bound, while unbound DNA fragments flow through. Filters are then washed extensively with binding buffer, and DNA fragments, retained on the filter by

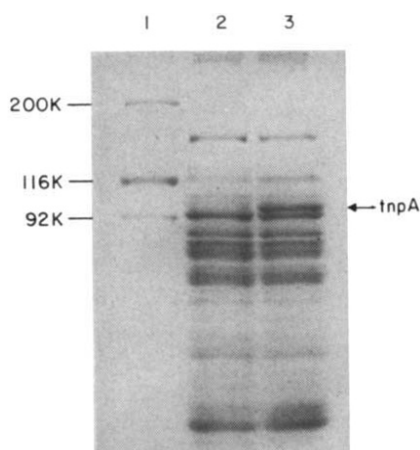


Fig. 1 Transposase protein accumulates in DS410 minicells. Lane 1 contains BioRad high- M_r markers as indicated. Proteins synthesized in DS410 cells containing pMB8::Tn3 72 (lane 2; transposase-negative)⁶ or pMB8::Tn3 5 (lane 3; transposase-positive)⁶ were separated on a 10% polyacrylamide gel and stained with Coomassie brilliant blue. The transposase band (tnpA) is clearly visible.

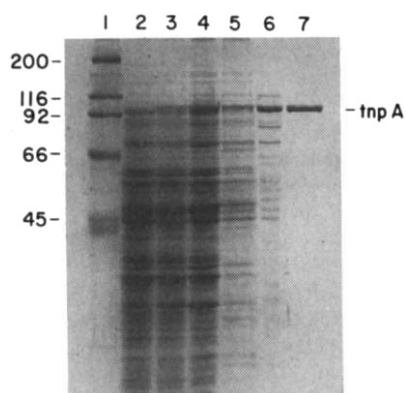


Fig. 2 Purification of the Tn3 transposase. Protein samples obtained at various steps in the purification scheme were separated on a 10% polyacrylamide gel and stained with Coomassie brilliant blue. Lane 1, BioRad high- M_r protein markers ($\times 10^{-3}$) as indicated; lane 2, total cell lysate from transposase-negative control strain pMB8::Tn3 72/DS410; lane 3, marker total cell lysate from transposase-overexpressing strain pMB8::Tn3 5/DS410; lane 4, total cell lysate from transposase-overproducing strain pMB8::Tn3 5/DS410 from which transposase was purified; lane 5, proteins extracted from membrane fraction with 2% Triton X-100; lane 6, nucleic acid binding proteins from the membrane fraction were precipitated by adding polyethyleneimine (final concentration 1.5%), the pellet was then extracted with 1 M ammonium sulphate, the extract was step-dialysed to remove ammonium sulphate, and proteins which precipitated in the 50–20 mM ammonium sulphate step were run in this lane; lane 7, transposase protein (4 μ g) eluted from a phosphocellulose column via a linear gradient of 50 mM to 1 M potassium phosphate.

binding to the transposase protein, are eluted in 1 ml elution buffer (0.2% SDS, 30 mM Tris-HCl pH 7.6), ethanol-precipitated, and run on polyacrylamide gels next to marker digestions of the plasmid from which the original DNA (pMB8::Tn3 digested with *Hpa*II) was obtained. The results of these experiments indicate whether or not the transposase is able to bind to DNA and also provide information about binding specificity, as DNA fragments retained by the transposase protein can be compared directly with the original DNA fragments.

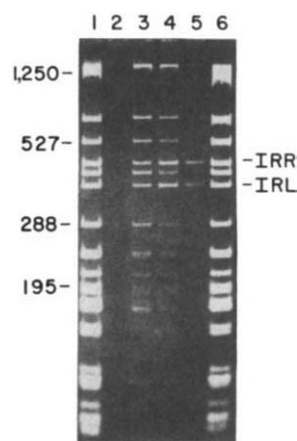


Fig. 3 Transposase binds to linear double-stranded DNA fragments. Lanes 1, 6, marker lanes of pMB8::Tn3 digested with *Hpa*II. Lanes 2–5, *Hpa*II fragments retained on nitrocellulose filters in the presence, respectively of 0 μ g transposase, 2.5 μ g transposase, 2.5 μ g transposase plus 8 mM ATP, and 2.5 μ g transposase plus 10 μ g ml⁻¹ heparin with 8 mM ATP. Representative fragment sizes are indicated in base pairs. The fragments containing the Tn3 left (IRL) and right (IRR) inverted repeats are also indicated.

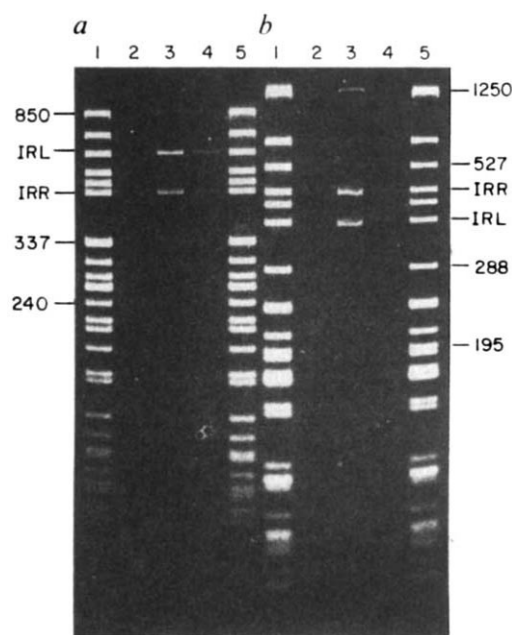


Fig. 4 Transposase site-specific DNA binding with DNA from *Cfo*I (a) and *Hpa*II (b) digestions of pMB8::Tn3; a, lanes 1, 5, *Cfo*I digestions of pMB8::Tn3 lanes 2–4, *Cfo*I DNA fragments bound in the presence of 0 μ g transposase, 2.5 μ g transposase with 8 mM ATP, and 2.5 μ g transposase with 16 mM ATP, respectively. b, Lanes 1, 5, *Hpa*II digestions of pMB8::Tn3, lanes 2–4, DNA fragments bound in conditions identical to those in a, lanes 2–4, respectively.

Because Tn3 moves as a non-permuted element, that is, the linear arrangement of the transposon is unchanged during transposition¹⁶, some factor must exist which is able to recognize the ends of the transposon as a first step in the transposition process. As the transposase is known¹⁷ to be required for transposition of Tn3 and as the transposase is known to act in *trans*^{2,17}, it is logical that such a site-specific binding activity capable of recognizing the ends of Tn3 might reside in the transposase protein.

To determine whether or not the transposase possessed the ability to bind specifically to the ends of the Tn3 transposon, we performed the above binding assay in a variety of conditions. Altering KCl, MgCl₂ and heparin concentrations affected the relative efficiency of binding, but no specific binding of DNA fragments containing the inverted repeats of Tn3 was observed. Altering binding time or protein/DNA ratios also had no effect on binding specificity (data not shown).

Because some of the enzymes involved in DNA replication and recombination have been shown to use ATP as a free energy source^{18,19}, we added ATP to our reaction mixture to determine whether ATP might affect the DNA binding activity of the transposase protein. We found that ATP produced a dramatic change in DNA binding activity; in the presence of 8 mM ATP, the transposase protein was able to bind specifically to DNA fragments containing the ends of the Tn3 transposon in a complex which was resistant to challenge with 10 µg ml⁻¹ heparin (Fig. 3, lanes 4, 5). In the absence of ATP, all DNA fragments could be removed from the transposase protein simply by adding heparin to 10 µg ml⁻¹, and no specific DNA binding was observed. In the presence of 8 mM ATP, fragments containing the 38-bp inverted terminal repeats of the Tn3 transposon remained bound to the transposase protein even in the presence of 10 µg ml⁻¹ heparin (Fig. 3, lane 5; Fig. 4b, lanes 1-5).

The specificity of this binding activity was confirmed by repeating the experiment using pMB8: Tn3 DNA digested with a different enzyme, *Cfo*I. Again, only the fragments containing the Tn3 inverted repeats were retained on filters in the presence of ATP and heparin (Fig. 4a, lanes 1-5).

Our results suggest that ATP qualitatively alters the binding specificity of the transposase protein. However, one alternative explanation is that more than one DNA binding site exists on the transposase protein, only one or a few of which are capable of binding to the ends of the Tn3 transposon. In this case, the ATP might not be required directly for recognition of the Tn3 inverted repeats, but might merely decrease the nonspecific DNA binding activity relative to site-specific DNA binding activity. However, this seems not to be the case because, at a heparin concentration of 10 µg ml⁻¹, in the absence of ATP, all DNA fragments are competitively removed and no binding specificity is observed.

It is possible that ATP addition alters metal concentrations, causing the transposase protein to bind specifically to the inverted repeats of Tn3. Preliminary results from experiments in which residual magnesium was removed by the addition of EDTA to 25 mM indicate that magnesium is not essential for specific binding of the transposase protein, although it does seem to increase the efficiency with which the transposase protein binds to the DNA fragments containing an inverted repeat.

As transposons promote non-homologous genomic rearrangements, they are believed to be evolutionarily significant. Our finding that the transposase is capable of binding to the ends of the Tn3 transposon and that this binding is specifically affected by ATP is exciting in that it suggests that ATP might be fuelling at least the initial step in the transposition mechanism (that is, recognition of the ends of the transposon by the transposase). We are now attempting to analyse the nature of this effect of ATP on binding specificity in greater detail. We are also investigating other transposition-related activities which current models²⁰⁻²² for the molecular mechanism of Tn3 transposition would predict to be associated with an active transposase or transposase complex.

We thank Dr Fred Heffron for providing the Tn3 mutants, Drs Marshal Elzinga and Nick Alonzo for the amino-acid analysis and protein sequencing, and Drs David Shortle, Don Oliver, Karen Armstrong, Robert Cabelli and Laura Liss for advice. W.L.W. was supported by NIH training grant CA09176.

4. Heffron, F., Bedinger, P., Champoux, J. J. & Falkow, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 702-706 (1977).
5. Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136**, 742-756 (1978).
6. Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. *Cell* **18**, 1153-1163 (1979).
7. Gill, R. E., Heffron, F. & Falkow, S. *Nature* **282**, 797-801 (1979).
8. Chou, J., Lemaux, P. G., Casadaban, M. J. & Cohen, S. N. *Nature* **282**, 801-806 (1979).
9. Dougan, G., Saul, M., Twigg, A., Gill, R. & Sherratt, D. *J. Bact.* **138**, 48-54 (1979).
10. Fennelwald, M. A., Gerrard, S. P., Chou, J., Casadaban, M. J. & Cozzarelli, N. R. *J. biol. Chem.* **256**, 4687-4690 (1981).
11. McCormick, M., Wishart, W. L., Ohtsubo, H., Heffron, F. & Ohtsubo, E. *Gene* **15**, 103-118 (1981).
12. Adler, H. I., Fisher, W. C., Cohen, A. & Hardigree, A. A. *Proc. natn. Acad. Sci. U.S.A.* **57**, 321-326 (1966).
13. Roosen, K. J., Fenwick, R. G. Jr & Curtis, R. III *J. Bact.* **107**, 21-33 (1971).
14. Ditto, M. D. *et al. J. Bact.* **149**, 407-410 (1982).
15. Wishart, W. L., Machida, C., Ohtsubo, H. & Ohtsubo, E. *Gene* **24**, 99-103 (1983).
16. Rubens, C., Heffron, F. & Falkow, S. *J. Bact.* **128**, 425-434 (1976).
17. Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136**, 742-756 (1978).
18. Weinstock, G. M., McEntee, K. & Lehman, I. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 126-130 (1979).
19. Gellert, M., Mizuuchi, K., O'Dea, M. H. & Nash, N. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3872-3876 (1976).
20. Grindley, N. D. F. & Sherratt, D. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1257-1261 (1979).
21. Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933-1937 (1979).
22. Galas, D. J. & Chandler, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4858-4862 (1981).

Domains of action of *bithorax* genes in *Drosophila* central nervous system

Erik Teugels & Alain Ghysen

Laboratoire de Génétique, Université Libre de Bruxelles, 67, rue des Chevaux, 1640 Rhode-St-Genèse, Belgium

The diversification of all *Drosophila* segments posterior to the mesothorax is controlled by a single gene cluster, the *bithorax* complex (BX-C)¹. Various observations have led to the conclusion that the BX-C is involved in the segmental determination of both the central nervous system (CNS) and the epidermis¹⁻⁶. In general, the neural and epidermal effects of a given BX-C mutation were found to occur in the same segment. However, only one neural structure was examined in each of these experiments; furthermore, the segmental origin of these few structures was in most cases unclear, making it impossible to decide whether or not the action of the BX-C genes is confined to the same regions in the epidermis and the CNS. Here we describe a set of 18 uniquely recognizable fascicles in the adult CNS, and establish the segmental origin of each of them. We analyse the effect of different BX-C mutations on these fascicles, and show that each mutation affects a region that overlaps two consecutive segments. We conclude that in the CNS, the spatial domains of action of BX-C genes are segment-length units which are out of register with the epidermal segments.

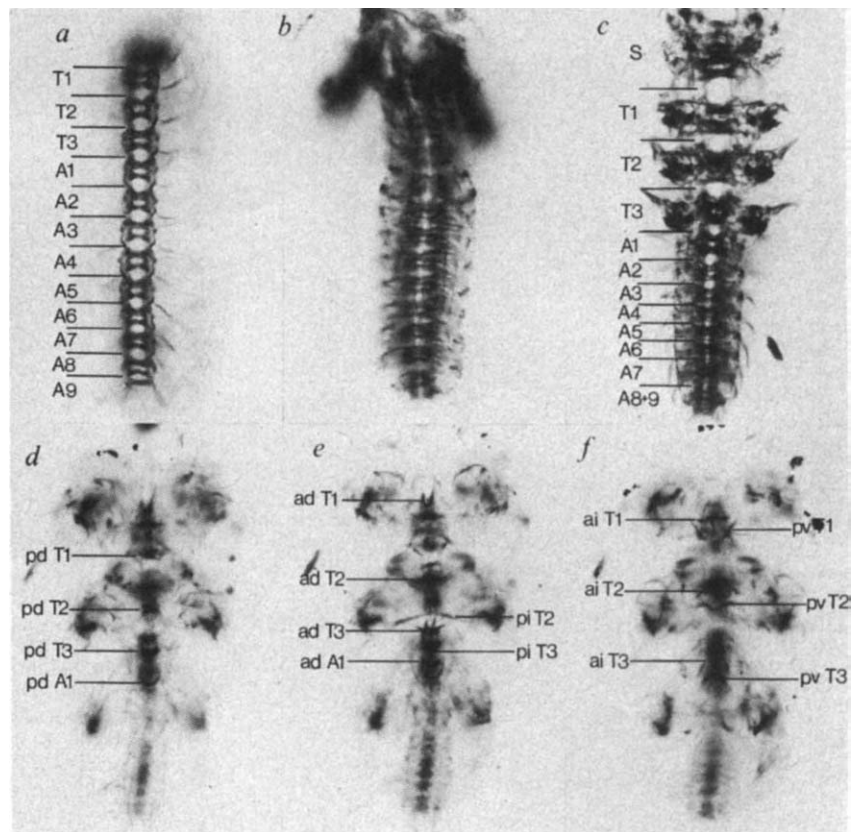
The problem of defining the spatial domains of action of BX-C genes in the adult CNS is complicated by two facts. First, there has been a tendency towards the condensation of the CNS during insect evolution. Whereas in primitive insects the segmental ganglia are clearly separated, in flies the thoracic and abdominal ganglia fuse to form a single ventral ganglion with no detectable intersegmental boundaries. Second, very few segment-specific structures have been identified in the CNS of *Drosophila*, and therefore the description of the segmental transformations associated with a given BX-C mutation is severely limited. We have overcome both difficulties by studying the pattern of transversal commissures in the CNS of adult flies, using a monoclonal antibody (5D12, isolated by Y. N. Jan and L. Y. Jan) which labels a set of transversal fascicles at all developmental stages (Fig. 1). Other labelled structures are also present in the adult CNS (Fig. 1d-f), but will not be dealt with here. Early during neurogenesis, two transversal commissures, one anterior and one posterior, are formed within each segmental ganglion, both in *Drosophila*⁷ and in the grasshopper⁸ (Fig. 1a). In the grasshopper^{8,9} and probably also in *Drosophila*¹⁰, these two commissures are pioneered by neurones belonging to the same segmental ganglion. By the end of embryonic development, a number of different fascicles can be distinguished in each of these two commissures by using antibody

Received 6 November 1984; accepted 17 January 1985.

1. Calos, M. P. & Miller, J. H. *Cell* **20**, 579-595 (1980).
2. Kleckner, N. A. *Rev. Genet.* **15**, 341-404 (1981).
3. Heffron, F. in *Mobile Genetic Elements* (ed. Shapiro, J.) 223-260 (Academic, New York, 1983).

Fig. 1 The development of transversal fascicles revealed by antibody 5D12. *a*, CNS of a 12-h-old embryo. One anterior and one posterior commissure are seen in the three thoracic (T1–T3) and eight abdominal (A1–A8) segments. The ninth abdominal ganglion (A9) contains only one commissure. The three suboesophageal ganglia and the brain stems are out of focus. *b*, CNS of an early first-instar larva. Various distinct fascicles, only some of which are in focus in this photograph, can be distinguished within each commissure. *c*, CNS of a 9-h-old pupa. The major changes that will lead to the adult pattern have only just begun and the suboesophageal ganglia are still closely apposed to the thoracic ganglia. The organization of the fascicles in two clusters, one anterior and one posterior, can still be detected. *d*, Adult CNS, dorsal focal plane. *e*, Adult CNS, intermediate focal plane. *f*, Adult CNS, ventral focal plane. Most of the adult fascicles studied in the mutants can be detected, although some are slightly out of focus (compare with Fig. 2, bottom row). Abbreviations: ad, anterior dorsal fascicle; ai, anterior intermediate fascicle; av, anterior ventral fascicle; pd, posterior dorsal fascicle; pi, posterior intermediate fascicle; pv, posterior ventral fascicle.

Method. In all cases, the staining was done on formaldehyde-fixed nervous systems except for *a*, where heptane/formaldehyde-fixed whole embryos were used. The dissection, fixation and rinse were performed in phosphate buffer 0.1 M, pH 7.6. All the following steps were each performed for 6 h in the same buffer containing 0.3% deoxycholate and 0.3% Triton (PDT buffer): The samples were incubated in antibody 5D12 (1:10 in PDT containing 10% of normal goat serum), rinsed in PDT, incubated with biotinylated secondary antibody (1:200 in PDT), rinsed in PDT, incubated in avidin biotinylated peroxidase (1:1:100 in PDT) and rinsed in PDT. For larval ganglia, the incubation times were reduced to 3 h instead of 6 h. All secondary immunoreagents were from Vector Labs. The samples were then processed for peroxidase staining, dehydrated, cleared and mounted. The adult ganglion was understained for photograph purposes.



5D12 (Fig. 1*b*). Little change is observed in this pattern up to the early pupal stage (Fig. 1*c*), but during the next 2 days most of the abdominal fascicles disappear and the thoracic fascicles undergo major modifications such that the segmental origins of the adult structures become less clear (Fig. 1*d–f*). However, by studying intermediate pupal stages, we have been able to follow the fate of a given fascicle throughout metamorphosis (Fig. 2) and have traced all the transversal fascicles that we detect in the adult CNS to the anterior or posterior embryonic commissure of a particular segment.

To simplify the description of mutant phenotypes, we have assigned a symbol to each adult fascicle within a given segment. The symbol summarizes the origin of the fascicle ('a' or 'p' for fascicles derived from the anterior and posterior embryonic commissures) and its position within the adult ganglion ('v', 'i' or 'd' for ventral, intermediate or dorsal). Some fascicles revealed by antibody 5D12 still show a clear homology from segment to segment in the adult, and will not be considered further. However, most of them have assumed segment-specific shapes in the adult (for example, compare the ad fascicles in segments T1, T2, T3 and A1; Fig. 2) thereby enabling us to determine the effect of BX-C mutations on this process of segmental diversification.

We have examined how this unique pattern of fascicles is affected by mutations in three consecutive regions of the BX-C: the *Ubx*, *bxd* and *iab-2* regions (Table 1, Fig. 3).

All recessive mutations in the *Ubx* region¹¹ produce transformation of the anterior metathorax (T3) into an anterior mesothorax (T2) in the epidermis. Two classes of mutations, *bx* and *abx*, have been distinguished on the basis of their localization, partial complementation and minute differences in

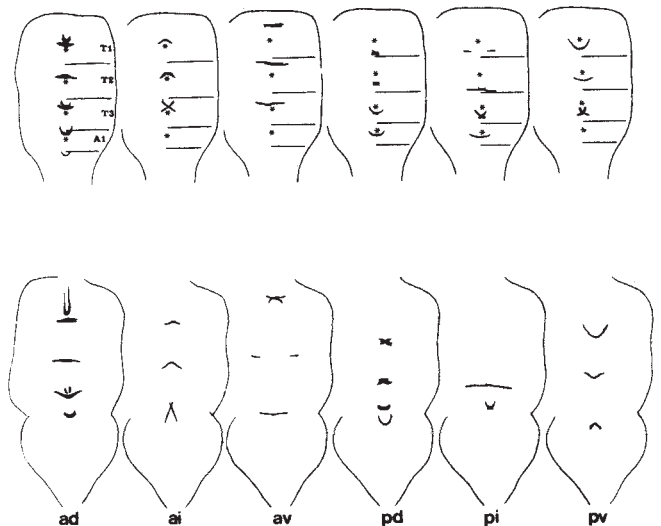


Fig. 2 Metamorphosis of the CNS. The transformation of the larval pattern into the adult pattern was followed at 9, 15, 20, 24, 48 and 73 h after the onset of metamorphosis. The first line of pictures are camera lucida drawings of a CNS during early metamorphosis (15 h APF); the second line shows drawings of the adult pattern. Each row indicates the segmental homologues for one of the fascicles studied in this paper. In the pupal CNS, the large gap between commissures belonging to consecutive segments is marked by a line; the smaller gap between the anterior and posterior commissures within a segment is marked by an asterisk. Abbreviations as in Fig. 1.

Table 1 Effect of BX-C mutations on adult *Drosophila* fascicles

	T2						T3						A1						A2					
	a	a	a	p	p	p	a	a	a	p	p	p	a	a	a	p	p	p	a	a	a	p	p	p
	d	i	v	d	i	v	d	i	v	d	i	v	d	i	v	d	i	v	d	i	v	d	i	v
<i>bx³</i>																								
<i>bx³/Df</i>																	+		+			+		
<i>bx^{34e}</i>							(±)																	
<i>abx¹</i>							+			+	+													
<i>abx¹/Df</i>							+			+	+						+		+			+		
<i>abx²/Df</i>							+			+	+	(±)					+		+			+		
<i>bx^d</i>										(±)	+	+	+	+	+	+								
<i>bx^d/Df</i>										(±)	+	+	+	+	+	+			+			+		
<i>Df/+</i>																	+		+			+		
<i>iab-2</i>																	+		+			+		
<i>pbx</i>																								
<i>pbx/Df</i>										(±)	+		+	+			+		+			+		
<i>bx^d¹/Df</i>										(±)	+		+	+			+		+			+		
Genotype							No. of nervous systems analysed																	
<i>bx³</i>																								
<i>bx³/bx³</i>							9																	
<i>bx³/Df(3R)bx^d¹⁰⁰</i>							4																	
<i>bx³/Df:</i>																								
<i>bx³/Df(3R)Ubx¹⁰⁹</i>							8																	
<i>bx^{34e}</i>																								
<i>bx^{34e}/Df(3R)bx^d¹⁰⁰</i>							7																	
<i>abx¹</i>																								
<i>abx¹/abx¹</i>							7																	
<i>abx¹/Df(3R)bx^d¹⁰⁰</i>							8																	
<i>abx¹/Df:</i>																								
<i>abx¹/Df(3R)Ubx¹⁰⁹</i>							13																	
<i>abx²/Df:</i>																								
<i>abx²/Df(3R)Ubx¹⁰⁹</i>							10																	
<i>bx^d</i>																								
<i>bx^d¹⁰⁰/bx^d¹²⁵</i>							11																	

The extent and frequency of the transformations are represented as follows. + Represents a rather complete transformation, +/− represents a partial transformation. In the latter case, either a fascicle is of an intermediate type or more often a normal and a transformed fascicle coexist. In one case (ai A1 in *pbx/Df*), +/− represents an additional fascicle that resembles not the adult ai T3, but its early pupal stage (see Fig. 2). Bold type indicates that the effect was seen in all mutants, normal type indicates an effect observed in a majority of the cases, and parentheses indicate an occasional effect. The results obtained with different genotypes were pooled whenever the results appeared to be identical. The lower half of the table shows the genotypes and numbers of nervous systems analysed.

phenotype¹². We examined four mutations in this *Ubx* region: two *bx* alleles, *bx³* and *bx^{34e}*, and two *abx* alleles, *abx¹* and *abx²*. None of the *bx* mutations gave any detectable effect on any fascicle, other than an occasional slight malformation of ad T3 in *bx^{34e}* hemizygous flies. This is perhaps not surprising in view of the fact that *bx* mutations, and in particular *bx³*, have been reported to have no effect on sensory neurones^{13,14}, and at most a minimal effect on the CNS^{15,16}. On the other hand, both *abx* alleles have a marked effect on one mesothoracic fascicle (pi T2) and on three metathoracic fascicles (ad, av and pd T3). These effects are detected in homozygous as well as hemizygous flies. In addition, a slight effect on pi T3 can be detected in some hemizygous *abx²* flies.

Two classes of recessive mutations are known in the *bx^d* region¹¹; *bx^d* mutations transform the anterior first abdominal segment (A1) into an anterior T3, and to a lesser extent the posterior T3 into a posterior T2, while *pbx* mutations show only the latter transformation¹⁷. We examined four mutants in the *bx^d-pbx* region: two extreme *bx^d* alleles, *bx^d¹⁰⁰* and *bx^d¹²⁵*, and the two existing *pbx* alleles, *pbx¹* and *pbx²*. Both *bx^d* mutations had a marked effect on two metathoracic fascicles (pi and pv T3) and on four first abdominal fascicles (ad, ai, av and pd A1), while a weaker effect was occasionally observed on pd T3. These effects were observed in heterozygous *bx^d¹⁰⁰/bx^d¹²⁵* as well as in hemizygous flies. The two *pbx* alleles showed essentially no phenotype when homozygous or when combined with a deficiency for the *Ubx* region (and part of the *bx^d* region), *Df(3R)bx^d¹⁰⁰*. However, they caused clear transfor-

mations when combined with a larger deficiency deleting the entire *Ubx-bx^d* region (and part of the *iab-2* region), *Df(3R)P10*. In this combination, *pbx* also affected both metathoracic (pi T3, and occasionally pd T3) and first abdominal (ad A1, and to a lesser extent ai A1) fascicles.

Extreme *iab-2* mutations, like *TP10*, are embryonic lethal when homozygous. Whether or not this boundary coincides with a compartment boundary cannot be decided, since it is not known if compartments exist in the CNS. However, the analysis of mutant embryos revealed that the anterior part of the second abdominal segment (A2) is transformed into A1 (ref. 18). In the CNS, however, transformations were observed even in *TP10/+* heterozygotes, indicating haplo-insufficiency for *iab-2*⁺. In this case, the mutant phenotype does not lie in the transformation of existing fascicles, but in the persistence of abdominal fascicles that are normally suppressed during metamorphosis. Here again, the effect is seen in two consecutive segments; an additional pi fascicle is usually present in A1, and additional ad and pd fascicles are often observed in A2.

We conclude that the defects corresponding respectively to the *abx*, *bx^d* and *iab-2* mutations define three consecutive regions in the CNS; their boundaries do not coincide with segmental boundaries, as in each case fascicles belonging to two different segments were affected. The linear order of these regions is entirely consistent with the observation made by Lewis¹ that consecutive genes in the BX-C are required in consecutive regions of the body, and indicates that the expression of the BX-C largely follows the same sequential

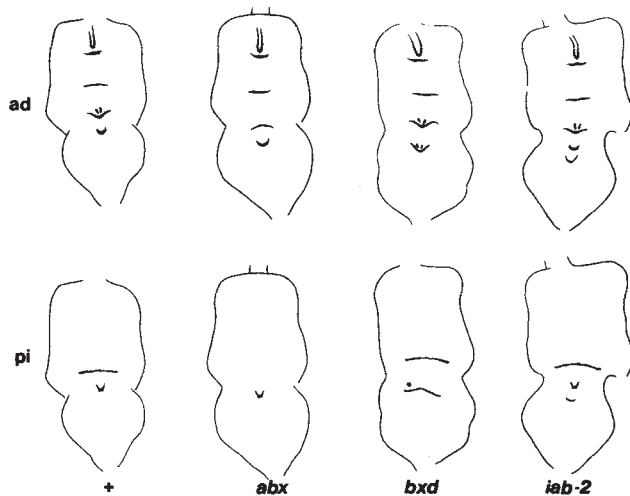


Fig. 3 Effect of BX-C mutants on the adult fascicles. The transformations undergone by two fascicles, an anterior one (ad) and a posterior one (pi), are illustrated for mutations in each of the three regions examined. *abx*: *abx*¹/*Df*(3*R*)/*bxd*¹⁰⁰ flies; *bxd*: *bxd*¹⁰⁰/*bxd*¹²⁵; *iab-2*: *T*(2;3)*P10*/+. In all cases, the modification results in a fascicle resembling that of the anterior segment. Thus, in some cases the shape of the fascicle is changed (for example pi T3 in *bxd*); in some cases the fascicle disappears (for example pi T2 in *abx*); and in some an additional fascicle is formed (for example, pi A1 in *iab-2*). All figures are camera lucida drawings; for each genotype, we used a ganglion which showed a complete transformation; see Table 1 for the frequency of the different transformations in each genotype.

strategy in the epidermis and the CNS. The fact that the boundaries of these regions do not coincide with segmental boundaries clearly supports the 'frameshift' hypothesis proposed by Lewis¹⁹ that the posterior regions of segments 'appear, in the presence of the mutants, to behave as though they were embryologically related to the anterior portion of the segment that follows, rather than precedes, them'. According to this hypothesis, the genes of the BX-C would act on segment-length units which include the posterior portion of one segment and the anterior portion of the following segment, and therefore are out of register with the segments themselves. In the case of the *Ubx*⁺ gene analysis of the distribution of *Ubx* proteins in the CNS has provided a direct confirmation of this hypothesis^{20,21}.

The epidermal segments consist of two compartments, one anterior and one posterior²². It has been hypothesized that compartments rather than segments are the basic units of development²³, and that the limits of the domains of action of BX-C genes coincide with compartment boundaries^{22,24,25}. Struhl²⁴ has recently rephrased the frameshift hypothesis in terms of compartments on the basis of his and other²⁶ results. Thus, the realm of action of BX-C genes would extend from an antero-posterior compartment boundary to the next one, and therefore include the posterior compartment of one segment and the anterior compartment of the following segment.

In the CNS, the three domains of action defined respectively by the *abx*, *bxd* and *iab-2* mutations extend from a posterior fascicle (pi) to another posterior fascicle (pd Table 1), implying that the limits of the domains are located in the posterior commissure. The observation that pd T3 is occasionally affected in *bxd* mutants suggests that pd may be at the very boundary between consecutive domains. Whether or not this boundary coincides with a compartment boundary cannot be decided, since it is not known if compartments exist in the CNS. However, the analysis of mutant phenotypes indicates that each transformation affects an entire segment-length unit. This result is particularly interesting in the phase of the *pbx* and *bxd* mutations, which in the adult epidermis affect predominantly the posterior T3 and anterior A1, respectively^{1,17,27}. Our results revealed no clear difference in the domain of action of *pbx* and *bxd*. Thus,

in different *pbx* combinations, whenever a transformation was detected, it affected fascicles on both sides of the T3-A1 segmental boundary. Similarly, when we reduced the *bxd* phenotype by using leaky *bxd* mutations the transformations in T3 and A1 were affected equally (Table 1). This indicates that a compartmental specificity of BX-C mutations similar to that observed in the epidermis is not readily detectable in the CNS. Thus, if compartments exist in the CNS, they do not seem to be important in the pattern of expression of BX-C genes.

We conclude that in the CNS, BX-C mutations affect entire segment-length units, suggesting that the domains of action of BX-C genes define 'genetic segments', similar to the morphological segments in their size but not in their boundaries. We emphasize, however, that our conclusions hold only for neurones involved in the diversification of the transversal fascicles labelled by 5D12; other structures may reveal a different pattern.

We thank Drs Y.N. and L. Y. Jan for allowing us to use their monoclonal antibody 5D12. We also thank E. Lewis for mutant strains, C. Goodman for discussion, P. Beachy and C. Dambly-Chaudière for improving the manuscript and R. Thomas for support. This work was supported by a FNRS research grant. E.T. is fellow of the IRSIA and A.G. is chercheur qualifié of the FNRS (Belgium).

Received 4 January; accepted 11 February 1985.

- Lewis, E. B. *Nature* **276**, 565-570 (1978).
- Jimenez, F. & Campos-Ortega, J. A. *Wilhelm Roux Arch. ges. Physiol.* **190**, 370-373 (1981).
- Green, S. H. *Nature* **292**, 152-154 (1981).
- Teugels, E. & Ghysen, A. *Nature* **304**, 440-442 (1983).
- Thomas, J. & Wyman, R. *Dev. Biol.* **102**, 531-533 (1984).
- Ghysen, A., Jan, L. Y. & Jan, Y. N. *Cell* (in the press).
- Jan, L. Y. & Jan, Y. N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2700-2704 (1982).
- Goodman, C. S., Bate, M. & Spitzer, N. J. *Neurosci.* **1**, 94-102 (1981).
- Raper, J. A., Bastiani, M. & Goodman, C. S. *J. Neurosci.* **3**, 31-41 (1983).
- Thomas, J. B., Bastiani, M. J., Bate, M. & Goodman, C. S. *Nature* **310**, 203-207 (1984).
- Bender, W. *et al. Science* **221**, 23-29 (1983).
- Lewis, E. B. in *Developmental Biology Using Purified Genes* (eds Brown, D. D. & Fox, C. F.) 189-208 (Academic, New York, 1981).
- Palka, J. & Schubiger, M. in *Development and Neurobiology of Drosophila* (eds Siddiqui, O., Babu, P., Hall, L. M. & Hall, J. C.) 223-246 (Plenum, New York, 1980).
- Ghysen, A., Janson, R. & Santamaria, P. *Dev. Biol.* **99**, 7-26 (1983).
- Ghysen, A. *Nature* **274**, 869-872 (1978).
- Strausfeld, N. J. & Singh, R. N. in *Development and Neurobiology of Drosophila* (Eds Siddiqui, O., Babu, P., Hall, L. M. & Hall, J. C.) 267-290 (Plenum, New York, 1980).
- Lewis, E. B. *Am. Nat.* **84**, 73-89 (1955).
- Morata, G., Botas, J., Kerridge, S. & Struhl, G. *J. Embryol. exp. Morph.* **78**, 319-341 (1983).
- Lewis, E. B. *Cold Spring Harb. Symp. quant. Biol.* **16**, 159-174 (1951).
- White, R. A. H. & Wilcox, M. *Cell* **39**, 163-171 (1984).
- Beachy, P. A., Helfand, S. L. & Hogness, D. S. *Nature* **314**, 545-551 (1985).
- Garcia-Bellido, A., Ripoll, P. & Morata, G. *Nature new Biol.* **245**, 251-253 (1973).
- Crick, F. H. C. & Lawrence, P. *Science* **189**, 340-347 (1975).
- Struhl, G. *Nature* **308**, 454-457 (1984).
- Lawrence, P. & Morata, G. *Cell* **35**, 595-601 (1983).
- Morata, G. & Kerridge, S. *Nature* **290**, 778-781 (1981).
- Minana, F. J. & Garcia-Bellido, A. *Wilhelm Roux Arch. ges. Physiol.* **191**, 331-334 (1982).

Erratum

Change in motion of Pacific plate at 5 Myr BP

A. Cox & D. Engebretson
Nature **313**, 472-474 (1985)

ON page 472, the last line of the fourth paragraph was deleted. The second sentence should read: "In offshore basins in central California, there is further tectonic evidence for a component of plate convergence across the San Andreas fault beginning at ~5.5 Myr⁴."

This should be followed by a new paragraph beginning: "Along the Pacific/Juan de Fuca (PA-JF) spreading centre, two major shifts in rotation poles occurred..."

Model for Grampian tract evolution

DEWEY and Shackleton's stimulating and provocative model for the evolution of the Grampian tract of the Caledonide–Appalachian orogen¹ postulates subduction towards the south-east between about 510 Myr BP and at least 470 Myr BP. As a result, they reject previous suggestions that the extensive layered mafic and ultramafic intrusions of Connemara and north-east Scotland were emplaced (at 485–500 Myr BP) above a north-west-dipping subduction zone and formed the roots of Andean-type volcanic arcs² or the lower parts of batholiths, subsequently removed during Upper Ordovician erosion³. The more mafic of these plutons contain cumulus assemblages of the minerals olivine plus chromite, followed by magnesian ortho- and clinopyroxene, very calcic plagioclase, and interstitial primary amphibole—inferred to precipitate at deep-crust levels within island arcs from subduction-related, silica-saturated hydrous magnesian magmas^{4,5}.

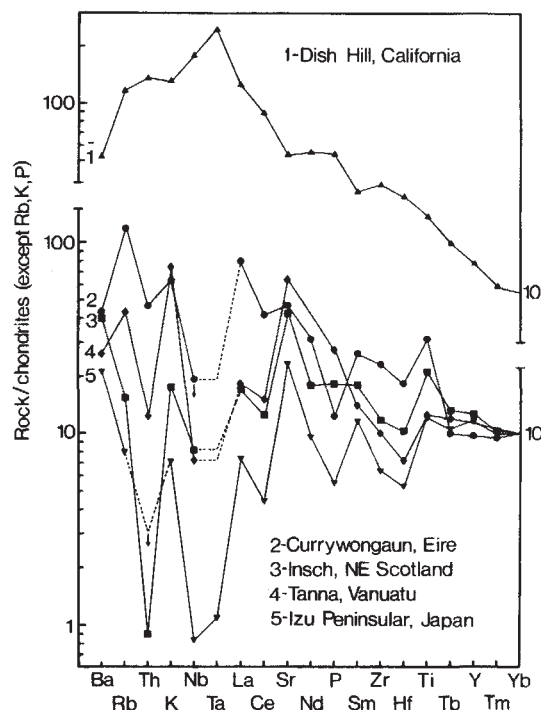
Figure 1 shows normalized incompatible trace elements in samples of each complex that have essentially escaped both the interaction with upper-crust metasediments and the Grampian deformation and metamorphism that affects most of them. Subduction-related basalts from Japan and Vanuatu are also plotted in Fig. 1. The essential similarities of many inter-element ratios between these modern circum-Pacific island-arc lavas and the Ordovician plutons are clear. As a group, the trace-element patterns (Fig. 1) of these samples are characteristic of subduction-related magmas worldwide and are quite different from those erupted in other tectonic environments, such as strike-slip pull-aparts, which Dewey and Shackleton suggest as an alternative. A Recent basanite from Dish Hill, Southern California (Fig. 1) is a typical example of the predominant type of basic magmatism in the south-west USA, extruded since the inception of a strike-slip plate boundary in this region.

Dewey and Shackleton have commented¹ that their "model is invalid, at least for the British Isles", if these Lower Ordovician plutons are subduction-related. I agree that more research is needed, but nevertheless, emphasize that there is already sufficient knowledge of the geochemistry of Lower Ordovician magmatism in the Caledonian–Appalachian orogen that the data from this approach cannot be ignored when constructing overall models.

R. N. THOMPSON

Department of Geology,
Imperial College,
London SW7 2BP, UK

Fig. 1 Normalized³ trace-element plot. Small arrows indicate element abundances below detection limits. The log scale represents: Abundance (mass) of each element in rock/Abundance in chondrites. Data are from refs 3 and 6.



1. Dewey, J. F. & Shackleton, R. M. *Nature* **312**, 115–121 (1984).
2. Yardley, B. W. D., Vine, F. J. & Baldwin, C. T. *J. geol. Soc. Lond.* **139**, 457–463 (1982).
3. Thompson, R. N., Morrison, M. A., Hendry, G. L. & Parry, S. J. *Phil. Trans. R. Soc. A310*, 549–590 (1984).
4. Leake, B. E. *Geol. J. spec. Publ.* **10**, 221–248 (1978).
5. Perfit, M. R., Gust, D. A., Bence, A. E., Arculus, R. J. & Taylor, S. R. *Chem. Geol.* **30**, 227–256 (1980).
6. Dupuy, C. et al. *Earth planet. Sci. Lett.* **60**, 207–225 (1982).

DEWEY AND SHACKLETON REPLY—

We thank Thompson for his geochemical data and conclusions and agree that a successful model for the evolution of the Grampian tract must account for geochemical and petrological data. Clearly, from the perspective of incompatible element pattern recognition, the Connemara and Banffshire intrusions are similar to the magmatic rocks of some modern arcs, but our lack of geochemical expertise disqualifies us from commenting on that data. However, we note that calc-alkaline magmatism occurs in a wide variety of non-arc settings including collisional environments such as eastern Anatolia. An important mechanism in collisional orogens is lithospheric delamination¹, which allows hot fertile mantle to gain access to levels as high as the base of a thickened crust. This causes rapid uplift and denudation and a surge of mafic magmatism and possibly partial melting of a mafic lower crust. We suggest that such a mechanism was responsible for the Grampian mafic intrusions. Our use of the term 'pull-apart' is simply kinematic and does not imply a particular regional tectonic setting. Pull-aparts in strike-slip systems occur in almost every type of regional tectonic environment including island

arcs. In collisional systems, they may localize calc-alkaline magmatic rocks (for example, Erzinçan basin on the North Anatolian Fault). A particular problem with the Connemara intrusions is that we know little of their original form and intrusive mechanism, though they appear to have a regionally disrupted sheet-like form. A particular problem in considering the intrusions as arc roots is that they were intruded into a pile of nappes during the latest stages of shortening. A West Pacific model, therefore, seems inappropriate because most of the West Pacific is in deviatoric tension where shortening is rare.

Lastly, the model is developed largely from data for Newfoundland where it seems difficult to refute, and then extrapolated to the British Isles. It is quite certain that the obduction model in its simplest form cannot account for the Taconian events of the Appalachians south of Quebec, because the tell-tale mafic detritus, particularly chromite, is virtually absent from the Taconic flysch belt in the US Appalachians. Hence, although it is also possible that the obduction model is inappropriate for the early Caledonides and is unique to the Canadian Appalachians, we cannot yet think of a better model.

J. F. DEWEY

R. M. SHACKLETON

Department of Geological Science,
University of Durham,
South Rd, Durham DM1 3LE, UK

1. Bird, P. *J. geophys. Res.* **83**, 4975–4987 (1978).

Trading places

Duncan Davies

Research and Innovation: A Record of the Wolfson Technological Projects Scheme 1968 - 1981.

By L. Rotherham.

Clarendon: 1984. Pp.99. £12.50, \$18.95.

Foresight in Science: Picking the Winners. By John Irvine and Ben R. Martin.

Frances Pinter, London/Longwood, Dover, NH: 1984. Pp.166. £12.50, \$21.

THOSE choosing the investments needed to generate technological developments with long lead times have at least three difficult fundamental realities to cope with.

First, the Top Nation (Spain four hundred years ago, France two hundred, Britain one hundred, the United States recently, Japan soon, perhaps China after that) has the easiest job; such a nation can afford to be involved in all important fields, and so does not have to face the politically awkward decisions about what *not* to do. Accordingly, its technology choices are concentrated on the more detailed matter of the picking and efficient prosecution of *project* winners. This is by no means easy, but it is an area where lobbying from those rejected is a little more manageable. Other, smaller economies have to contract out of some important topics because they cannot afford everything, and thus have to begin with criticism from the lobbyists for the items that lose out. This is true whether the rejection is explicit, courageous and rational, or whether it is left to the shorter-term working of the market process.

Secondly, there are two kinds of forecast necessary for these decisions, whether or not they are prepared explicitly. On the one hand, one needs to predict the general climate of social relevance and acceptance for possible new technologies. On the other hand, one has to foresee the relative positions and comparative advantages of different competitors, whether nations or sub-national institutions such as firms or research teams. For the first, one is predicting an environment over which one has little control; for the second, one is in the business of partly self-fulfilling prophecy. A team having cohesion and enthusiasm can, with a less attractive technology, sometimes prevail over another with better propositions but which is bedevilled by internal rivalries and inefficiencies. (This point is forcefully made by Irvine and Martin on p. 144 of their book, where they point out that self-fulfilling prophecy is an important component of Japanese success.) So the two parts of policymaking — trying to predict the weather, and deciding between umbrellas, parasols and other devices — must be to some extent separate.

Thirdly, in retrospect no forecast can be seen to have been absolutely right; even though one protagonist predicts so much better than everyone else that his nation or

firm gains great comparative advantage, he will have been wrong in many conspicuous respects and practical success will usually be accompanied by unjust criticism of (often irrelevant) error. This is why the ambitious get into politics and stay out of explicit forecasting if they possibly can. "Act promptly, but never give your reasons; your actions may well be right, but your reasons will usually be wrong." "Treason doth never prosper; what's the reason? Why, if it prosper, none dare call it treason."

Against this background, it is not surprising that books on the conduct and success of forecasting in science and technology make melancholy reading, especially in those countries that once were Top Nation and are now no longer. Of the two books reviewed here, *Foresight in Science* (which should really have been called *Foresight in Technology*) is concerned with national technology policies and decision-making at the macro-level, and only incidentally covers the science that underpins them, while Dr Rotherham's book, *Research and Innovation*, is a record of the Wolfson Technological Projects Scheme which, between 1968 and 1981, allocated a total of £17 million to 186 projects — an average of just under £100,000 each. But, because such micro-decisions determine whether the bright people in universities languish unfunded, devote themselves entirely to intellectual curiosity, or seek to understand and link their work with social needs, the importance of the Wolfson scheme is greater than the size of the sums might suggest.

Many funds finance projects and study; the Wolfson initiative, master-minded by Lord Zuckerman, sought to break new ground by using money to launch teams that could thereafter earn their own living by trading in the technology that they generated. Grants were usually to be once-and-for-all, so that minds were to be concentrated by the prospect of being hanged. Thus, the assessors of the many applications were going much further than the normal process of peer judgement on scientific merit, and were attempting to forecast business success. By comparison with the rest of the academic funding system in Britain, the Wolfson scheme has had, *pro rata*, a good deal of impact on the economy at large. It must, however, be recognized

that the Wolfson teams usually drew heavily on the results of work funded by the British Research Councils or University Grants Committee. Thus, Wolfson has sought to set up a series of do-it-yourself "Micro Research Development Corporations" alongside the National Research Development Corporation. The original objective was that each team should survive after the end of the grant by selling its efforts and wares in the market, and to generate capability for innovations rather than for measurement and more orthodox consultancy services. However, this objective has been broadened, for Wolfson has now endorsed as "success" two other kinds of team self-sufficiency.

One is that of those groups which, though not yet trading successfully in the market, have enhanced their thrust and esteem enough to win support from the Research Councils at the end of the Wolfson grant; the other is the achievement of the groups which have met quite new needs in measurement and technical consultancy. This point is illustrated by Dr Rotherham's thumbnail sketches of each of the projects funded, as well as his overall critique, which show that the total success rate of over 50% includes about 10-20% as "innovation contribution", about 20% as "contract measurement and consultancy" and about 20% funded after Wolfson from the public purse. These are impressive results, and we really need a more precise assessment along these lines. Dr Rotherham does not attempt it, and indeed is reflective and descriptive rather than judgemental, possibly on the reasonable ground that comparisons are divisive and that academic dog should try to avoid eating academic dog. His book is nonetheless a most useful quarry for information about the Wolfson units and ideas about adventurous funding methods.

Another comparison can be made between the numbers of awards to different institutions and of resultant commercial successes. Four universities were awarded around £1 million or more: Imperial College, Manchester, Southampton and Birmingham. Of these, Southampton has the biggest yield of market trading, but mainly in the assessment, advice and measurement business. It is, however, rather early to draw firm conclusions about improved methods for project choice. Perhaps it is not unexpected that the pattern of Wolfson esteem for different universities, as measured by grant levels, is very different from that of the University Grants Committee as shown by the pattern of the 1981 cuts in their funding.

Foresight in Science, which was written under contract to the UK Advisory Committee for Applied Research and Development (ACARD), is concerned almost entirely with the machinery used for technical forecasting in different countries, and the notice taken of such forecasts and agencies by the political masters. There is good

justification for the lack of attention the authors give to case detail in that there is fair unanimity in different countries' choice of the technological areas likely to be of most importance during the next few decades. Consequently, most interest lies in the extent to which the forecasts are used, taking national capabilities and disabilities into account in determining policy (if any) in this area.

On this basis, France and Germany emerge as confused and unimpressive; the United States pluralistic and somewhat wasteful; Japan mysterious, purposeful and effective; and the private sector everywhere assiduous but rather blinkered. The snag is that formal forecasting is always supplemented by unofficial, less transparent procedures. In France, for example, the polytechniciens do not seem to pay much attention to the machinery for science funding, and the key decisions — for example those concerning the reconstruction of the auto industry after the collapse of Citroën, the building of a common engine plant, the methods for ensuring the use of French computer equipment, and the early entry of France into microelectronic telephone switching through the system E10 — are not closely linked with the academic scientists and their work. The polytechniciens carefully watched the Colloque Nationale, and the grand Gallic gestures of M. Chevènement. They were then able to cash in on any beneficial trends when they were brought back to clear up the chaos.

In Germany, it seems to be the engineers rather than the scientists whose priorities have most effect on business decisions and are most responsive to them. Again, there are informal alliances which play a large part in German industrial technology but which are not part of official machinery. In the United States, public purchasing of prototypes and funding of private sector research through NASA and the Department of Defense moves very large sums of money in such a way as to lengthen the forward-looking of American firms. However, the policies of President Reagan, especially those currently designed to cut the budget deficit, are eroding energy and other technologies with long lead times. In all three countries, as well as Britain, strong internal rivalries and conflicts have diminished the extent to which choices, once made, are pursued loyally and single-mindedly. France, Germany, the United States and Britain have often failed to enter the recommended and important area of self-fulfilling prophecy.

It is in Japan that formal and informal processes operate hand-in-hand, thereby ensuring that decisions get the maximum chance to work; for example, the development of ceramic gas sensor systems has been helped by a law compelling every house to have a gas leak detector. Japanese weakness in primary, counter-cultural innovation is offset by the prompt identification of successful new ideas from other

countries, and early inward licensing while it is cheap. Japanese technologists have surveyed the world, and brought foreign teams to Japan, to encourage competitiveness in their next round of computer and control-engineering development. Perhaps more important still, because of the large proportion of the gross national product that is invested, the Japanese banks are able to insist on long-term technical planning as a condition for loans. Although a very great deal happens that is not visible on the surface, the unity of purpose makes the visible moves a better guide to events and prospects than is the case elsewhere.

What will not be seen for a long time is whether the elaborate, dedicated and expensive Japanese effort being devoted to the science base will bear fruit, or whether social dedication to teamwork and co-operation will inhibit the idiosyncratic innovator. Real originality — the eccentric thinking of the counter-cultural heretic — is not understood or tolerated by a disciplined social machine such as that in Japan today.

Irvine and Martin's book is a valuable source of information on models and systems used in various countries, and how they are meant to work. Confusion, however, arises from their intermingling of attempts to generate policy for technology on the one hand, and science on the other. It is easier to identify technological priorities, and to plan accordingly, than to plan science. Irvine and Martin say clearly that market-pull and ideas-push are needed in collaboration, but the technology does not usually arise sequentially from science (despite the understandable belief to the contrary in the science community); usually, it uses science as a source of strength and reinforcement. Some science planning can be done in response to technological opportunity and need, but this must not exclude curiosity-orientated work, which formerly was cheap enough to be based on peer judgement of the promise of brilliant individuals. Unfortunately, the growing expense of some curiosity-based science is now creating a very awkward need to make choices on the basis of such difficult concepts as "number of major surprises per megabuck". This deeply passionate process must be kept separate from technological planning, or muddle ensues. □

Duncan Davies, formerly Chief Scientist and Engineer at the Department of Industry, and Research and Development General Manager of ICI, is a consultant and Chairman of the ceramic development company BCRA Ltd.

● Longman has recently published a collection of essays examining the implications for scientific endeavour of the increasing links between academic research and industry. The book is entitled *Science as a Commodity: Threats to the Open Community of Scholars*. The editors are Michael Gibbons and Björn Wittrock, price is £9.95 in paperback.

Mathematics and membranes

W.D. Stein

The Physical Chemistry of Membranes.

By M.E. Starzak.

Academic: 1984. Pp.334. \$47, £38.

Equations of Membrane Biophysics.

By N. Lakshminarayanaiah.

Academic: 1984. Pp.426. \$69, £55.50.

READERS of *Nature* will be aware of the present excitement in membrane biophysics. Our ability, by the patch-clamp techniques of Neher and Sakmann, to measure the electrical properties of single ionic channels in membranes means that the molecular events in channel opening and closing, and in ionic movement through channels, can be monitored. More and more ionic channels are being isolated biochemically, the genes coding for the channel proteins cloned and sequenced, and the channel proteins themselves investigated by physical techniques. Structural models of such channels have been proposed, and the theoretical investigation of ion movement through channels is at a stage where the electrical properties of channels will soon be predictable from them. The study of membrane carriers and pumps lags behind, but pumps have been crystallized in two-dimensional arrays, and the structure of a pump solved, albeit at a low resolution. The dynamics of the membranes themselves are beginning to be understood, the viscosities of regions of cell membranes measured, and the intra-membrane and trans-membrane diffusion of membrane components determined and understood in physicochemical terms.

All this is the meat of university courses on membrane biophysics. I know, however, from personal experience, how difficult it is to convey the sense of new excitement while still providing the firm basis of physical chemistry that students need for a critical appreciation of the modern approach. The subject is of its nature heavily mathematical and quantitative. What, then, is one to do with this mathematics? Write it all out on the blackboard — or merely refer the student to the original (often difficult) paper? Both of the books under review try to solve this problem by providing a text which includes many of the relevant equations. Unfortunately, neither book is really satisfactory.

Starzak's book suffers in part from its misleading title. One would expect that *The Physical Chemistry of Membranes* would treat the membrane's physical chemistry — the forces that mould and stabilize it, its viscosity, the diffusion within the membrane of phospholipids and proteins. These topics are not in fact covered, and the emphasis falls on the surface chemistry

of the membrane and on ion movement across it. The problem, it seems, is that Starzak has felt obliged to deal with a number of general questions of solution chemistry, elementary and irreversible thermodynamics, chemical and enzyme kinetics — even electronics! — on all of which topics excellent books are available at both elementary and advanced levels. Even the relevant topics that he does include, such as ion movements across excitable membranes, are well discussed, at the same level, in other texts. A further defect is the exceptionally poor bibliography. No attempt is made to refer the student to available texts on the subjects covered, and of the 30-odd papers quoted nearly half are citations (sometimes incorrect) of sources from which diagrams have been taken.

Lakshminarayanaiah's book is a more substantial effort. It is comprehensive in scope, covering most of what one would expect to find on ion transport across membranes, with treatment (more elementary)

also of non-electrolyte flows. It is hard going in parts, especially where the meaning of a symbol alters over a few pages. I found the treatment of surface potentials quite opaque as these changed names (from distribution to outer voltage potentials, from Lange to contact or from inner to Galvani) and interchanged symbols (from Ψ to Φ to V). The bibliography again is unsatisfactory, being perfunctory and even irrelevant in the earlier chapters. But the main drawback of the book is its style. There are difficulties with the language itself — such as the absence of the definite article — and obscurities in the text from which it is inordinately difficult to wrestle meaning.

There was certainly a place in the literature for both of these books. However, in order successfully to fill this place they should, perhaps, have been more keenly criticized while still in manuscript. □

W.D. Stein is Professor of Biophysics at the Hebrew University, Jerusalem.

Vision interrupted

John Mollon

Visual Masking: An Integrative Approach.

By Bruno G. Breitmeyer.

Clarendon: 1984. Pp.454. £25, \$34.95.

MAN's resolution in time is poor. If we are to analyse two successive stimuli separately, or if we are reliably to report their order, then the interval between them must exceed 50-100 milliseconds. Our modern machines have a much better resolution; and at least one specialized subsystem in the human brain — the subsystem that discriminates the directions of sounds — enjoys a temporal resolution that is almost four orders finer than that of the system as a whole.

The coarseness of our temporal resolution is well illustrated by "backward visual masking". If a target letter is presented for only one millisecond, it will be easily identified by a human observer, provided it is of sufficient contrast; but if it is followed within, say, 50 milliseconds by a spatially superimposed montage of letter fragments, then it will be masked. Especially curious is the form of masking called "metaccontrast": if a bright square patch is presented briefly and is followed within 50 milliseconds by two flanking patches that are of the same luminance as the first patch, then our perception of the first patch is almost completely suppressed.

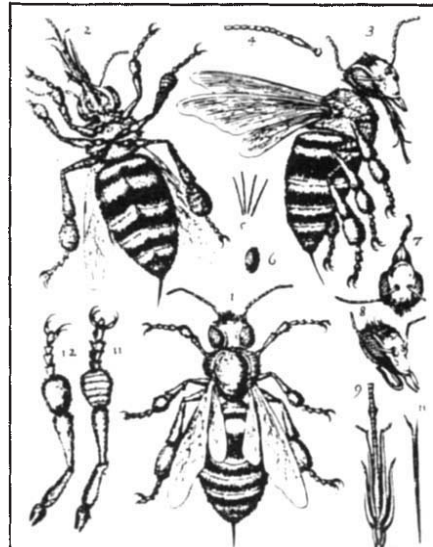
Such phenomena promise a means to analyse the early processes of visual perception; and therefore masking has been experimentally studied for over a century. There has not previously been a book on the subject, and so it is valuable to have Dr Breitmeyer's comprehensive survey.

Among the topics he covers are persistence of vision, masking by flash and by pattern, metaccontrast, the effects of stimulation of the peripheral visual field, and the integration of information from successive fixations. One of Breitmeyer's strengths is his excellent grasp of the classical German literature on masking. He does, however, neglect the French literature, and particularly wanting is discussion of the papers of Jeanine Blanc-Garin. Nor is there any reference to Baylor and Hodgkin's work on vertebrate photoreceptors, a weird omission in a book concerned with the temporal aspects of vision.

Breitmeyer's skill lies in giving accurate short accounts of experiments and theories. However, he could have been more critical of many of the studies included. For example, he devotes several pages to an analysis by Michaels and Turvey that purports to show that the first and second of two stimuli contribute equally to the formation of a "cortical icon". The analysis depends on subtracting the masking curve for the non-dominant eye from that for the dominant eye; in each case, the ordinate is the percentage of trials on which the target is correctly identified and the abscissa is the temporal relationship of target and mask. The fallacy here is the common one of taking behavioural performance as a direct measure of some internal process or state. Breitmeyer unjustifiably assumes that the performance difference between, say, 100% and 60% correct corresponds to the same degree of internal integration as does the difference between 40% and 0%. (If this type of error had a recognized name, it might occur less often. "Fallacy of reification" would capture the essence of it.)

But worst of all, Breitmeyer is silent about an experimental scandal that has marred this field of research. Between 1965

and 1980, many experimenters (including Breitmeyer) uncritically used commercial tachistoscopes that had physical time constants much longer than manufacturers claimed. What happened was that the manufacturers — and those few experimenters who did bother to calibrate their tachistoscope lamps — used photocells sensitive to the fast, short-wave component of the lamp output whereas the eye was maximally sensitive to the slow, long-wave, lamp phosphor. The properties of commercial tachistoscopes were documented in 1978 (*Q. J. exp. Psychol.* 30, 555) and it would have been good to see Breitmeyer assess the problem as he reviewed the



*Nature magnified — enlarged views of the honey-bee, the earliest known studies made with the aid of a microscope, from Francesco Stelluti's *Descrizione dell'Ape* (1625). The illustration is reproduced from *Single Lens*, by Brian J. Ford, which will be reviewed in Nature's Spring Books issue on 25 April.*

literature. Since 1978, tachistoscopes have been largely replaced by computer-driven displays. Astonishingly, however, the persistence of the long-wave component of display phosphors has led to a fresh generation of experimental artefacts. Thus Chapter 10 of the book places great weight on experiments on integration across saccades, but these results had been shown to be artefactual before the date of Breitmeyer's preface (see, for example, *Science* 222, 188; 1983).

Breitmeyer finally gives himself away in his long "Epilogue", a rambling, overblown piece of bull in which Aristotle, Galileo, Descartes, La Mettrie, Kant, Goethe, Santayana, Bergson, Monod, Tolman, Lorenz, Piaget, Polyani, Popper, Dawkins, Breitmeyer and others, jostle promiscuously. It is troubling that the publisher did not have the editorial firmness to delete this embarrassing 20-page essay. □

John Mollon is a Lecturer in Experimental Psychology at the University of Cambridge.

Good vibrations

Colin Gough

The Physics of the Violin.

By Lothar Cremer.

MIT Press: 1985. Pp. 450. \$35, £41.95.

FOR over 150 years scientists have been investigating the physics of the violin. Some have been motivated by the hope of discovering the "Stradivarius Secret" — the scientific explanation of the highly-valued tone quality of many instruments of the famous Cremonese violin makers of the seventeenth and eighteenth centuries. For others, such as the author of this book, the attraction has been to understand the acoustical and vibrational physics that ancient luthiers empirically incorporated into their instrument designs, culminating in the "modern" violin around 1550.

In the nineteenth century, musical acoustics was at the forefront of science. People of the stature of Helmholtz and Rayleigh, often through their interest in acoustics, were laying the wave-mechanical foundation of modern physics. And in the early part of this century, Raman, later to win the Nobel prize for his work on optical spectroscopy, published his first scientific papers — on the physics of the violin. But thereafter, with the advent of the exciting new fields of quantum science, musical acoustics became something of a backwater.

In recent years, however, the subject has experienced something of a revival, arising partly from the recognition of a number of interesting problems of "classical" physics still to be solved, and partly because modern experimental and computational techniques now allow a much more realistic modelling of the science involved. Prominent in such activities have been Professor Cremer and his co-workers at the Institute of Acoustics in Berlin. Over the past 25 years, this group has published a succession of important papers on musical acoustics. Now Professor Cremer has summarized that work and has brought together material that was hitherto widely dispersed in the literature. Although no attempt has been made to provide a fully comprehensive coverage of current research, generous space is devoted to closely related work of other authors.

Cremer treats the violin simply as a mechanical structure that is forced into vibration by the bowed motion of the string. Almost half of the 450 pages are devoted to the bowed string problem, reflecting the author's main research interest. The transformation of vibrational energy by the supporting bridge, now widely recognized to be a major factor in determining the tone of an instrument, is discussed in detail, as are the vibrational modes of the instrument as a whole. This logical approach to the energy chain is concluded with a discussion of sound

radiation and the way that this sound is modified by the environment.

Professor Cremer hopes his book will appeal not only to the scientist but also to the violinist or violin-maker wishing to understand the underlying science of their instrument. Unfortunately, few musicians or craftsmen are likely to possess the mathematical background to cope with the relatively sophisticated theoretical description of the complex vibrational characteristics of the violin. On the other hand, the book should certainly convince any scientific reader that musical acoustics, and the acoustics of the violin in particular, is a perfectly respectable and fascinating field of applied science; indeed recent research on the violin has led to a number of new ideas that could usefully be applied to many other vibrating systems.

Cremer makes no attempt to relate physical measurement to the subjective assessment of a violin's quality. As he explains in the preface, although modern science can provide a fairly accurate description of the physics involved in producing sound, we simply do not have the necessary language even to start to try and correlate them. This is a challenge which will probably keep physicists, psychoacousticians, players and instrument-makers busy for at least another 150 years.

Colin Gough is Senior Lecturer in the Department of Physics, University of Birmingham.

Power and penguins

David J. Rose

A Guide to Nuclear Power Technology: A Resource for Decision Making.

By Frank J. Rahn *et al.*

Wiley: 1984. Pp. 1,000. £101.30, \$106.

NUCLEAR steam supply systems have millions of parts, far more than Boeing 747s and in their way just as high technology. Describing this technology in a single book, albeit a big one, is a daunting task — some might say impossible. But it is the subtitle, *A Resource for Decision Making*, that is the reason for my concern about this book, of which more anon.

What is there here, in 500,000 words of small print plus 900 figures and tables? In 20 chapters, from basic processes, through radiation, the fuel cycle, reactors, materials, operation, safety, health and environment, regulation, economics, proliferation and intermediate stops, the book describes a particular world of nuclear power, often in more detail than we want. It is mainly an American world, of diagrams, photographs and words about all these subjects. Not wholly, of course, because some topics such as hydrogen flammability limits and radioactive toxicities know no national boundary. But the Canadian CANDU reactor gets a mere

eight pages, the British Advanced Gas Reactor four pages and the French reactor standardization one 10-word clause on page 799. Comment on the Far East, where many innovations are taking place, seems to be missing completely.

Even the great specificity about equipment and practices in the United States is not very well matched to the stated task of informing decision-makers. The book describes in great depth how things are now, somewhat mechanically: properties of steel, operation of the latches on a control rod drive, patterns of refuelling a light water reactor, an extractor evaporator for removing water from rod waste and a myriad of other things. Most of this, like the mythical book about penguins, gives the decision-maker more detail than he wants. It isn't wrong, but is it useful?

Decision-makers want to know other things. Why did the system work (or not) before? What are the present options? What are the trends? Compared to what? Here are a few examples.

- Why do nuclear power plants cost so much, when the entire nuclear steam supply system plus the turbine-generator costs only about 25% of the total? And why such large disparities from place to place? The chapter on economics lists increases in material costs, effect of lead time and so on, but still leaves a fuzzy impression.

- What about the prospect of substantially improved light water reactors for the near term? The work proceeds now, principally in Japan, aiming for higher capacity factor, less radiation exposure to workers, easier maintenance, more compatible materials and so forth. But I find nothing of this, only some four pages on such topics as the classic spectral shift idea. If decision-makers in the United States don't watch out, the initiative will pass from them, surely something they should know.

- The potential of smaller reactors is becoming an increasingly lively topic these days. Is there the possibility of overcoming the diseconomy of small size by the economies of serial, factory or modular construction? Such smaller reactors would decrease the financial exposure of electric utility companies. Again, there is nothing on the subject here.

Still, the book has its merits and can be useful, though not principally as a handbook for decision-making, let alone policy decisions or technology assessment. It is descriptive and not overly technical (as its preface says), and fairly well written. Of the twelve groups to whom the authors commend the book, two seem most likely to profit from it: engineers with an interest in but no special training in nuclear technology, and nuclear engineers wishing to get a broader grasp of certain aspects of nuclear technology. □

David J. Rose is Emeritus Professor in the Department of Nuclear Engineering at the Massachusetts Institute of Technology.